

IMPROVING THE SKILLS OF LOW-PERFORMING READERS IN AN
ALTERNATIVE SCHOOL PROGRAM

A Dissertation

by

ELIZABETH BUBONIC OLSON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: School Psychology

© 2004

ELIZABETH BUBONIC OLSON

ALL RIGHTS RESERVED

IMPROVING THE SKILLS OF LOW-PERFORMING READERS IN AN

ALTERNATIVE SCHOOL PROGRAM

A Dissertation

by

ELIZABETH BUBONIC OLSON

Submitted to the Office of Graduate Studies of

Texas A&M University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Jan E. Hasbrouck
(Co-Chair of Committee)

Douglas Palmer
(Co-Chair of Committee)

Richard I. Parker
(Member)

Malatesha Joshi
(Member)

Douglas Palmer
(Head of Department)

August 2004

Major Subject: School Psychology

ABSTRACT

Improving the Skills of Low-Performing Readers in an Alternative School
Program. (August 2004)

Elizabeth Bubonic Olson, B.S., University of Iowa

Co-Chairs of Advisory Committee: Dr. Jan E. Hasbrouck
Dr. Douglas Palmer

Research has demonstrated that many children and adolescents exhibiting behavior problems also evidence serious reading problems as well as a low self-efficacy (i.e., belief in their ability) toward reading. The consequences of these problems on both the student (e.g., dropping out of school) and society as a whole (e.g., cost to taxpayers) are serious and, in most cases, preventable. In order to prevent students from dropping out of school and to empower teachers with a method for removing disruptive students from the classroom, many states have implemented alternative education programs. The purpose of this study was to implement an effective reading intervention in a disciplinary alternative school where students were assigned from 20-40 days for infractions such as fighting, threatening others, and excessive office referrals. The design consisted of a series of 26 single-case AB studies. Subjects were ages 12-16 in a mid-sized city in Central Texas. There were 19 males and 7 females. Subjects were mostly of African American and Hispanic backgrounds, and the majority received some form of special education services. Data were analyzed using visual and statistical single case model techniques. Results suggest that an intensive oral reading fluency program can positively impact the oral reading fluency, accuracy, comprehension, self-efficacy toward reading,

and social comparison with regard to the reading ability of students placed in a disciplinary alternative education program on a short-term basis.

DEDICATION

To my husband Tim. You have always supported and encouraged me. You have always reassured me that I can accomplish anything I set my mind to. I know that you have put some of your dreams on hold so that I could achieve mine and for that I will always be grateful.

To my dad. From the beginning you have always told me that I can do anything I want. You gently guided and never pushed. Even though you knew it would be challenging, you never seemed to doubt that I would achieve my goals. I have had you as a role model as I see you treat others with dignity and respect, never caring about material gain as much as having a positive work ethic and high moral standards. Thank you for always believing in me.

To my mom. I watched you raise six kids, attend college at night, and work full-time. You did not complain and smiled through it all. When it gets difficult, I think of you and remember that it is possible to have it all.

ACKNOWLEDGMENTS

I would not have been able to accomplish my goals without the help of the following people. First, my mentor Jan Hasbrouck. I appreciate you fielding many questions and encouraging me throughout the process. Your advice is always treasured and I look forward to continuing to learn from you as a colleague. Thank you also to Richard Parker. I learned so much from you as an employee and a student. You helped inspire this dissertation and challenged me to learn more. Douglas Palmer, I was fortunate to have you as a teacher and co-chair. Thank you for your patience and ideas.

Laura Crozier, Roxanne Garza, Kristen Gsanger, Cassandra Romine, and Richard Evans put in many hours helping me test students. I had many things happening at one time and I would not have been able to complete this study without their assistance.

Thank you also to the Center for Collaborative Learning Communities at Texas A&M University for a generous grant in the amount of \$3,000, which enabled me to implement the study.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
DEDICATION.....	v
ACKNOWLEDGMENTS.....	vi
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
CHAPTER	
I INTRODUCTION.....	1
Problem Statement.....	1
Significance of the Problem.....	1
Reading.....	3
Self-Efficacy.....	4
Alternative Education Placement.....	5
Intervention.....	6
Current Study.....	6
Purpose.....	6
Research Questions.....	7
II LITERATURE REVIEW.....	8
Alternative Education Placement.....	8
Link Between Behavior and Reading.....	12
Reading.....	14
Stages.....	14
Accuracy, Fluency, and Comprehension.....	16
Curriculum-Based Measurement.....	18
Self-Efficacy.....	20
Social Comparison and Reading Performance.....	24
Read Naturally.....	27
Current Study.....	29
III METHOD.....	31
Subjects.....	31
Design.....	32
Instruments.....	34

CHAPTER	Page
Pretest Measures.....	34
Weekly Monitoring.....	36
Posttest Measures.....	38
Intervention.....	38
Procedure.....	39
Generalizability.....	40
Statistical Analysis.....	41
Does a Short-Term, Intensive Reading Intervention Targeting Oral Reading Fluency Improve the Reading Fluency and Accuracy Skills of Students Temporarily Assigned to An Alternative School Due to Behavior Problems?.....	41
Does the Reading Intervention Have an Impact on Reading Comprehension of the Students?.....	45
How Does This Intervention Affect the Students' Self-Efficacy With Regards to Reading?.....	46
How Does the Intervention Affect the Way Students Compare Themselves to Others With Regards to Reading?.....	46
IV RESULTS.....	47
Descriptive Statistics.....	48
Proximal Measure: Curriculum-Based Measurement.....	62
Visual Analysis.....	62
Statistical Analysis.....	65
Accuracy Skills.....	75
Cross-Correlation.....	83
Distal Measures.....	84
Does the reading intervention have an impact on reading comprehension of the students?.....	87
How does this intervention affect the students' self-efficacy with regards to reading?.....	88
Supplementary Informal Analysis.....	89
V DISCUSSION AND SUMMARY.....	93
Does a Short-Term, Intensive Reading Intervention Targeting Oral Reading Fluency Improve the Reading Accuracy and Reading Fluency Skills of Students Temporarily Assigned to an Alternative School Due to Behavior Problems?.....	94
Does the Reading Intervention Have an Impact on Reading Comprehension of the Students?.....	96
How Does This Intervention Affect the Students' Self-Efficacy With Regards to Reading?.....	98
How Does the Intervention Affect the Way Students Compare Themselves to Others With Regards to Reading?.....	99

Summary and Implications for Future Research.....	100
REFERENCES.....	104
APPENDIX A.....	113
APPENDIX B.....	116
APPENDIX C.....	126
VITA.....	136

LIST OF TABLES

TABLE		Page
1	Means and standard deviations of age, grade, reading level, and days stayed.....	48
2	Effect sizes, pre- and post-cleansing of fluency for 26 students using Simple Mean Difference model with 85% confidence intervals and standard errors.....	68
3	Effect sizes of fluency for 26 students using Mean + Trend Difference model with 85% confidence intervals and standard errors.....	69
4	Effect sizes of change between phases for fluency of 26 students using the Center Mean + Trend model with 85% confidence intervals and standard errors.....	73
5	Effect sizes pre- and post-cleansing of accuracy for 26 students using Simple Mean Difference model with 85% confidence intervals.....	77
6	Effect sizes and significance levels of accuracy for 26 students using Mean + Trend Difference model with 85% confidence intervals.....	79
7	Effect sizes of accuracy for 26 students using the Center Mean + Trend model with 85% confidence intervals and standard errors.....	80
8	Means and standard deviations for subscales measuring fluency and accuracy.....	85
9	Qualitative information and effect sizes using the Mean + Trend technique for each student.....	90

LIST OF FIGURES

FIGURE		Page
1	Schematic figure of study design.....	33
2	Power curve using multiple regression.....	45
3	Violin plot of Test of Word Reading Efficiency Sight Word subtest (mean = 100, SD = 15).....	49
4	Violin plot of the Test of Word Reading Efficiency Phonetic Decoding subtest.....	50
5	Violin plot of the Test of Word Reading Efficiency Total Word Reading Efficiency subtest.....	51
6	Violin plot of the Woodcock Reading Mastery Test—Revised Word Attack subtest.....	52
7	Violin plot of the Woodcock Reading Mastery Test—Revised Passage Comprehension subtest.....	53
8	Violin plot of the Gray Oral Reading Tests Third Edition Accuracy subtest.....	54
9	Violin plot of the Gray Oral Reading Test Third Edition Rate subtest.....	55
10	Violin plot of the Gray Oral Reading Test Third Edition Passage Comprehension subtest.....	56
11	Violin plot of the Gray Oral Reading Quotient Third Edition Oral Reading Quotient.....	57
12	Violin plot of the pretest measures, excluding the individual subtests of the Gray Oral Reading Tests Third Edition.....	58
13	Violin plot of the Gray Oral Reading Tests Third Edition subtests.....	59
14	Violin plot of the baseline (Phase A) and intervention (Phase B) phases for average curriculum-based measurement scores for 26 students.....	61

FIGURE	Page
15 Fluency graphs depicting a low variability, or bounce (left), and high variability (right).....	63
16 Graph of oral reading fluency illustrating improvement in mean levels between phases based on visual analysis.....	64
17 Graphs of oral reading fluency depicting no change in mean level between phases (left), and negative mean level changes (right).....	65
18 Examples of fluency graphs using the Simple Mean Difference and Mean + Trend techniques for Student G.....	71
19 Multiple regression graph of the effect sizes resulting from computation of time series data for the fluency measurement with 85% confidence limits using Center Mean + Trend.....	74
20 Accuracy graphs depicting low variability (left) and high variability (right).....	76
21 Multiple regression graph of the effect sizes resulting from computation of time series data for the accuracy measurement with 85% confidence limits, using Center Mean + Trend.....	82
22 Violin plot of cross-correlations (lag-1) of absentees with fluency and accuracy scores.....	84

CHAPTER I

INTRODUCTION

Problem Statement

National estimates indicate that approximately 30% of students are at great risk of academic failure (Halsell Miranda & Santos de Barona, 1990). According to information from the 1990 national census, approximately 5% of the 4,253,507 students in Texas public schools are considered at-risk for academic failure (Information Services, 2000). In addition, over 4,000 students in the state of Texas are assigned to temporarily attend school at an alternative education placement due to discipline problems in their home schools.

Significance of the Problem

It has been estimated that approximately 40% of all students will drop out of school (Horn, 1987). These numbers do not include students who have been expelled from school for offenses ranging from violent acts to repeated discipline problems. There are several risk factors that increase the likelihood of students dropping out of school, including academic failure, undiagnosed learning disabilities, frequent discipline referrals, developmental immaturity, emotional problems, physical disabilities, language problems, second language acquisition issues, socioeconomic status, and grade repetition (Committee for Economic Development, 1987; Horn, 1987). Although it has been hypothesized that students with lower cognitive ability evidence more behavior

This dissertation follows the style and format of *School Psychology Review*.

problems than those with higher cognitive functioning (Rutter, Tizard, & Whitmore, 1970), Stanton, Feehan, McGee, and Silva (1990) found that reading scores were a mediating factor, concluding that it is the student's reading level that is linked to behavior, not innate cognitive ability in and of itself. It appears that inattention and hyperactivity are strong indicators of academic problems for children. However, antisocial behavior and delinquency are stronger predictors of academic difficulty for adolescents (Hinshaw, 1992). A study conducted in a Missouri public school district and reported by Freeman, Gum, and Blackburn (1999) showed that a staggering 90% of the students that dropped out of school were either failing or experienced academic difficulty when they were in the 7th grade. Results such as these help to make strong arguments in favor of early intervention for at-risk students.

Although school failure has obvious consequences such as risk of poverty and limited job opportunities for individual students (Blau, 1981), society as a whole suffers as well. Catterall (1985) estimates that those students who drop out of school in the United States in any single year cost society \$240 billion in lost earnings and taxes over the course of their lives. Research has shown that low academic achievement, poor vocabulary skills, and weak verbal reasoning abilities by the end of elementary school accounts for 27% of the variance among delinquent students (Farrington, 1979). This is important because not only are school dropouts costly to society, delinquents cost taxpayers a considerable amount of money in court costs, paying for imprisonment, and direct costs of crimes such as vandalism and theft. Hale (2001) reported that

incarcerating a person for one year costs at least three times what it would cost to educate them during that time period.

There are clearly many factors contributing to low academic achievement and dropping out of school. Although some of the risk factors are beyond the control of school personnel (e.g., the student's socioeconomic status), there are ways in which schools can intervene and improve the academic development of at-risk students. One powerful step to take is to explicitly address a student's reading deficit. (Duckenfield, 1990; Reitzammer, 1991).

Reading

Researchers (e.g., LaBerge & Samuels, 1974; Samuels, 1986; Sanders, 2001) agree that reading consists of both decoding (i.e., deciphering how the word is pronounced) and comprehension (i.e., understanding the meaning of the word). The main debate has centered around delivery of instruction: Whether whole language methods (i.e., immersing students in reading for experience and motivation rather than teaching specific discrete skills) or phonics instruction (i.e., teaching rules and applied strategies for letter-sound correspondence) is superior for teaching children to read (Sanders, 2001). After much discussion and research, a general consensus has been reached that (a) it is crucial that students learn to decode words accurately through phonics instruction (National Research Council, 1998; Samuels & Flor, 1997), and (b) it is important to teach students to read in the context of connected text (Samuels, 1986) because gleaning meaning from words is the purpose of reading (National Reading Panel, 2000). Samuels likens this to advice often given regarding diet: Balance is best.

Another idea with which researchers tend to agree is that automaticity is crucial for successful reading. There are several different definitions of automaticity (e.g., Samuels & Flor, 1997; Sanders, 2001), but in general automaticity means the ability to decode with enough speed and accuracy that the action is automatic. LaBerge and Samuels (1974) theorized that once decoding is automatic for students, more cognitive attention may be allotted for comprehension. Automaticity (or fluency) is extremely important because if readers struggle with accuracy and rate, they have trouble comprehending what they read (Reutzel & Hollingsworth, 1993) and begin to lose interest in reading as well as related activities (Hasbrouck, Ihnot, & Rogers, 1999). According to Stanovich (1986), this lack of motivation can result in a downward spiral as students begin to read less and less for pleasure and fall further and further behind their peers in reading ability. Therefore, it would stand to reason that addressing students' reading deficiencies by systematically targeting reading fluency should lead to improvements in reading and comprehension and may indirectly have a positive impact on behavior.

Self-Efficacy

Bandura (1997) defined self-efficacy as "beliefs in one's capabilities to organize and execute the course of action required to produce given attainments" (p. 3). Put simply, if people have high self-efficacy with regard to a particular skill such as reading, they believe they have the ability to read well. Studies have shown that students' engagement in reading (Henk & Melnick, 1998) and reading achievement (Shell, Colvin, & Bruning, 1995) are impacted, whether positively or negatively, by their perceptions about their reading ability. Students who experience difficulty reading often become

frustrated and evidence “learned helplessness” (i.e., experiencing failure over and over until they believe they do not have the ability to read well) (Seligman & Maier, 1967). This may lead to antisocial behaviors in class to avoid having to read, or to mask the emotional trauma of failure (Hinshaw & Anderson, 1996). Therefore, it is important to investigate whether an intense, relatively short-term reading intervention can have a positive impact on self-efficacy related to student’s reading achievement.

Related to self-efficacy is social comparison (i.e., how people compare themselves to others). Researchers have found that people are most influenced by those whose performances are the most similar to theirs (Festinger, 1954; Wood, 1989). If Person A perceives Person B to be similar to himself in terms of ability to read, and Person B performs well on a reading measure, Person A will be more confident in his ability to perform well on the same measure. However, if Person B does poorly, Person A will feel less confident. Based on this relationship, it is likely that students will rate their ability to read based on how they perceive themselves in relation to their peers.

Alternative Education Placement

Because many students evidencing behavior problems may also have undiagnosed reading difficulties (Curtis & Longo, 1999) and because the alternative school setting allows for more individualized programming, personnel in such settings are in a unique position to intervene with these troubled students to provide the assistance they require to improve their reading skills. However, while the research base on effective reading interventions continues to grow, little or no research has been conducted with this population to guide development of programs to address this critical

area of need. Research that does exist is often focused on alternative programs that students elect to attend and are usually focused on drop-out prevention (Lange & Sletten, 2002). Research is often anecdotal (Southwest Educational Development Laboratory, 1995), usually focusing on processes rather than outcomes (Raywid, 1994). In addition, there are no state mandates requiring assessment and remediation of students' reading skills at alternative education programs using research-supported programs. Therefore, interventions vary widely between schools and there is no known documentation of the types of interventions implemented by alternative education placements.

Intervention

Read Naturally is a strategy that was developed by Candyce Ihnot, an elementary school teacher, to increase the reading fluency of her students in special education and remedial classes (Hasbrouck, et al., 1999). The program was developed based on an extensive review of the reading literature, with considerations such as the importance of repetition and feedback, self-efficacy, and necessity of goals in mind. Although the main focus of Read Naturally is to increase students' oral reading fluency, a brief comprehension component is included.

Current Study

Purpose. The purpose of the present study was to implement and evaluate the effectiveness of an intensive, short-term instructional intervention for reading difficulties in at-risk/disabled readers temporarily assigned to an alternative education setting. The project was designed to (a) improve the reading skills and self-efficacy (i.e., belief in their ability to read) in these students and (b) extend research into a previously unstudied

area: Effective reading interventions in alternative school programs with high-risk adolescents with antisocial externalizing behaviors.

Research questions. The purpose of the present study was to address the following research questions: (a) Does a short-term, intensive reading intervention targeting oral reading fluency improve the reading accuracy and reading fluency skills of students temporarily assigned to an alternative school due to behavior problems? (b) Does the reading intervention have an impact on reading comprehension of the students? (c) How does this intervention affect the students' self-efficacy with regards to reading? and (d) How does the intervention affect the way students compare themselves to others with regards to reading?

CHAPTER II

LITERATURE REVIEW

Alternative Education Placement

Although alternative education settings have existed in the United States since the 1960s (Young, 1990), in the early 1990s, many states began to see a need for alternative education programs for two main purposes: 1) to help students become productive members of society and 2) to empower teachers and schools to remove disruptive students from mainstream classrooms (e.g., expulsion, suspension), without leaving them with the only option of roaming the streets (Southwest Educational Development Laboratory [SEDL], 1995). Throughout the United States, there are many types of alternative education, ranging from optional (e.g., magnet schools, drop-out prevention programs) to mandatory placements (e.g., disciplinary alternative education placements) (Lange & Sletten, 2002), and on- or off-campus settings (McCreight, 1999).

The definition of alternative education varies from state to state and within communities, making an overall summarization of alternative education in the United States very difficult (Lange & Sletten, 2002). In some states, such as Michigan, schools are not required to establish alternative education settings. Other states mandate alternative education programs for “at-risk” students. Several papers have been written to describe programs in the state of Texas. In 1995, the state of Texas adopted a policy stating that each school district was required to set up an alternative education setting for students who were having behavior problems. The policy included the following as

criteria for placement: committing a felony, committing several specified serious offenses either at school or during a school-sponsored activity, or other behaviors specified in school districts' codes of conduct (Cortez & Robledo Montecel, 1999). The purpose of this legislation was to provide an environment in which students could receive individualized attention and support designed to guide them back onto the "right" path both behaviorally and academically.

According to Cortez and Robledo Montecel (1999), during the 1996-1997 school year in the state of Texas, 72,997 students were removed from their home campuses and placed in alternative education settings. Of these, almost three-fourths were due to violating the schools' code of conduct, rather than for legal offenses. Length of stay varied from an average of 16 days to more than 45 days. It is unclear from the data how many students were placed in the alternative setting more than once. In Texas, most students in disciplinary alternative education settings were minorities, with Hispanics overrepresented compared to the percentage of Hispanics that make up the general student population. Students of low socio-economic status (SES) and those in special education were also likely to be placed in these settings. These researchers also found a high rate of absenteeism that increased with the number of days assigned. Because the only data that is required to be collected from school districts is name, student identification number, date of birth, district type, reason for referral, and length of stay, we do not know how many students have reading or other academic difficulties, the typical reading level of each student, or what academic interventions have typically been attempted prior to the students' referral.

According to Cortez and Robledo Montecel (1999, p.5), “For the most part, alternative education programs are being used as dumping grounds for ‘undesirable’ students who, once there, get little support”. Although the legislation had the primary purposes of removing disruptive students from the classroom and empowering teachers with a method of discipline, the policy did not specify what was to become of students once they were placed in the alternative setting. In addition, according to Cortez and Robledo Montecel, there is no existing data as to the effectiveness of these programs. They state that, although several school districts provide anecdotal data as to their effectiveness, quantitative data is virtually nonexistent. This statement is supported by a review of the literature. SEDL (1995) presented a summary of research on alternative schools. Its conclusions were the same as Cortez and Robledo Montecel: “A closer look at the literature reveals an abundance of anecdotal evidence regarding the success of alternative education programs, but a dearth of well-designed, scientific research” (p.2). Research that claims alternative education programs raise students academic scores and improves their behavior usually refers to programs in which students choose their placement. No research studies were found that quantitatively supported disciplinary alternative education programs for improving academics and behavior of students. In fact, as stated by McCreight (1999), “So far, research on disciplinary programs show no positive long term gains and may even increase negative outcomes” (p.12). SEDL (1995) suggests that one reason for the lack of success by disciplinary programs is the short duration of most programs. In addition, although a large number of students spend time in alternative education placements every year in the state of Texas, there is not a

required research-supported behavior and/or academic remediation program, accountability system, or specialized certification requirement for staff members working with these at-risk students. There is also no requirement that students' progress be tracked once they transition back into their home schools.

Although studies have been conducted examining (a) how students in alternative programs viewed the flexibility and choices they had in the programming, (b) impact on self-esteem, and (c) academic achievement, most of these studies were conducted with students who chose to attend their alternative placement (Lange & Sletten, 2002). These schools typically were GED completion programs and schools designed for drop-out prevention. In addition, the schools studied varied in length of stay. Studies were usually process rather than outcome-driven (Raywid, 1994). No outcome-driven studies specific to short-term, disciplinary alternative education placements were discovered. While some researchers have attempted to conduct studies identifying best practices of alternative education programs, low return rates of surveys (e.g., 39% for a survey in Texas by McCreight, 1999) and lack of requirements to keep detailed data suggest that we do not know what the majority of disciplinary alternative education settings are providing in the way of academic programming, placement in reading levels, and support to students. Therefore, we do not know the potential of these settings to positively intervene with students' academic and behavior difficulties to provide the catalyst that can change the trajectory of the students' academic careers for the better.

Research has linked reading and student behavior problems (e.g., Hinshaw, 1992; Samuels, 1986). It is unclear if reading problems lead to behavior problems or vice versa

and this issue has been addressed at length in the literature. Students who experience difficulty reading often become frustrated and evidence “learned helplessness” (i.e., experiencing failure over and over until they believe they do not have the ability to read proficiently) and give up, which may lead to acting out behaviors in class (Hinshaw & Anderson, 1996). Research has also linked academic problems to grade retention and frequent discipline such as suspensions/expulsions, implying that behavior problems may cause academic problems (Hinshaw & Anderson, 1996). Despite this, as was eloquently stated by Curtis and Longo (1999), “it is not necessary to identify the underlying causes of a student’s reading problem in order to help” (p. 6).

The purpose of this study was to implement and evaluate the effectiveness of an intensive, short-term instructional intervention for reading difficulties in at-risk/disabled readers temporarily assigned to an alternative education setting. The link between reading, behavior, and self-efficacy toward reading is important because students in this study were assigned to an off-campus disciplinary alternative education setting due to behavior problems at their home campuses including physical fighting, excessive office referrals for refusing to follow teacher directives, ingestion of alcohol or drugs on campus, and threatening harm such as bringing a bomb to school.

Link Between Behavior and Reading

There is a large amount of literature documenting the link between behavior problems and academic difficulty. Dishion, Loeber, Southamer-Loeber, and Patterson (1984) studied 70 white male adolescents (approximately 32.9% of whom were currently or previously involved with law enforcement) in Oregon. Among the factors studied

were interpersonal problem solving skills, interpersonal competence, academic competence, reading achievement, verbal intelligence, homework completion, and chores at home for which each student was responsible. The researchers measured delinquency by studying official police records as well as self-reports of delinquent behavior during interviews and on the Delinquent Lifestyle Scale. The results of the study showed that those adolescents who had contact with police for non-traffic-related incidents did less homework at home, had poorer interpersonal problem solving skills, and demonstrated skill deficits in reading, school competence, and verbal intelligence.

In a study of 177 boys aged 7-12, most of whom qualified for diagnosis of a disruptive behavior disorder, Frick, et al. (1991) used a structured interview, strict diagnostic criteria, and standardized intelligence and achievement assessments to study the link between academic deficits and disruptive behavior disorders (i.e., Attention-Deficit/Hyperactivity Disorder [ADHD] and Conduct Disorder [CD]). Their findings confirmed the results of previous studies that found significant links between ADHD and CD and academic underachievement. However, they note that CD may have only been linked due to a high comorbidity with ADHD.

It has also been hypothesized that students with lower intelligence scores evidence more behavior problems. Therefore, it would stand to reason that these students would probably also have lower reading achievement scores. Stanton, Feehan, McGee, and Silva (1990) investigated the relationship between IQ, reading scores, and behavior problems. They studied the same cohort of children at two-year increments beginning at age 3, conducting the final assessment at age 15. Their study included measures of

family adversity, behavior problems, reading ability, and IQ using questionnaires and standardized assessments. They found that removing the effects of family adversity, problem behaviors early on, and IQ resulted in a significant association between reading ability and behavior problems. Their results indicated that upon entry to school, low IQ predicted problematic behaviors; however, reading ability predicted successive changes in behavior problems.

Hinshaw (1992) reviewed existing literature to summarize and clarify the link between reading deficits and behavior problems such as inattention, hyperactivity, and aggression. He included descriptions of studies showing that it was not intelligence that was most predictive of delinquent behavior. Rather, achievement deficits (reading achievement in particular) were the key predictors. Hinshaw concluded that inattention and hyperactivity were more closely linked to academic difficulty in childhood than was aggression, but in adolescence it was delinquency and antisocial behavior that were the strongest predictors of academic underachievement. Language delay at 3 years of age and an adverse family climate predicted reading deficits in early elementary school; however, social class and lower intelligence were not as predictive.

Reading

Stages. In trying to understand the process of reading and the deficits of a particular student, is useful to describe reading as a series of stages that each student undergoes. Based on numerous research studies, Chall (1983) asserted that normally-developing students undergo a series of stages in the development of reading skills. Although success in each stage is necessary for progression to the next stage, the skills

acquired in each stage do not stop developing once a student moves into the next stage. According to Chall, Stage 1 typically occurs around age 6-7 or grade 1-2. In this stage, which she terms the Initial Reading or Decoding Stage, students learn the letters of the alphabet as well as their sounds. In addition, children learn that there is a specific purpose for these letters—to put them together to make words.

Stage 2 occurs when the student is 7-8 years old or in grade 2-3. Chall describes this stage as confirmation, fluency, ungluing from print. In this stage, the students are able to combine the lessons learned in the first stage and increase their level of fluency (i.e., rate and accuracy). Chall points out that the primary purpose of this stage is not to gain new information, but to strengthen the skills learned in the first stage. In Stage 2, students ideally gain more confidence in their reading. Chall references studies by Kraus (1973) and Bloom (1964), which showed that Stage 2 appears to be a crucial stage; if students achieve significantly below same-age peers during this stage, they will experience academic failure during their school years unless they receive help. During this stage, it is critical for students to practice reading and be exposed to many books.

If students are not successful during Stage 2, they will have a difficult time during the next stage, labeled by Chall as Reading for Learning the New: A First Step. Stage 3 normally occurs at approximately age 9 or grade 4. During this stage, students already possess most of the requisite skills of reading and will use reading to obtain new knowledge. The complexity of this knowledge will increase as the student's background knowledge, vocabulary, and cognitive abilities develop and increase. Chall points out that it is at this time that public schools typically introduce subject areas to students (e.g.,

history, science, social studies). According to Chall, although students are beginning to read for meaning, their abilities are still very limited. Therefore, students in Stage 3 are still unable to take multiple viewpoints and are reading mostly to learn facts or how to do things.

Stage 4 is termed by Chall as Multiple Viewpoints and occurs during high school or when students are approximately age 14-18. As the label implies, the main element of this stage is that students are able to take multiple perspectives. In addition, the depth of material that student are able to read at this stage is much greater than what they have been exposed to before.

The final stage is Stage 5, Construction and Reconstruction—a World View. This stage occurs at approximately age 18 and above, or college age. During this stage, students are able to use highly developed skills profitably such as skimming the important information and choosing relevant information. In addition to abstracting relevant information, students at this stage are able to synthesize relevant information. Clearly there are students who do not move through these stages in a developmentally predicted manner. Students who struggle with reading can get “stuck” at any of these stages and need targeted interventions to improve their reading skills and move to the next stage.

Accuracy, fluency, and comprehension. Much of the research literature regarding reading focuses on students’ accuracy, fluency, and comprehension. According to LaBerge and Samuels (1974), reading is a highly complex skill and is hindered when too much attention is allotted to the process of reading. The researchers hypothesized that

many things can be processed in one's mind at a time, but only if attention is not required of more than one thing at a time. Therefore, automatic information processing, or automaticity, is crucial for developing reading skills. This is because if students' attention is required for decoding text, they will not have attention left to comprehend what they are reading. LaBerge and Samuels stated that one can determine if a skill has become automatic by investigating whether it can be completed (or processed) while one's attention is directed somewhere else. The keys to developing reading automaticity are practice, repetition, feedback that is specifically related to a criterion baseline, achievement of accuracy, distributed practice, and motivation (LaBerge & Samuels, 1974; Samuels & Flor, 1997).

Automatic, or fluent, reading is important not only because allotting too much attention to reading makes it a tedious process, but because it allows attention to be freed up to comprehend the meanings of the words on the page. If students must spend a lot of effort and attention to accurately associate a letter to its sound, then they will have a difficult time blending the sounds together to make a word. If the students have already allotted most of their attention to sounding out, or decoding, a word then it is unlikely that they will be able to hold the word long enough in short-term memory to retrieve its meaning. LaBerge and Samuels used this argument to emphasize the inadequacy of simply developing reading skills to accuracy—fluency is a crucial component.

Reutzel and Hollingsworth (1993) tested LaBerge and Samuel's theory regarding the connection between fluency and comprehension in a study involving 78 second graders. They assigned the student to either an experimental group in which they

received oral reading fluency training using a model, practice, perform technique, or a control group in which students simply took turns reading aloud. Instruction lasted for four months. Students were administered the Iowa Test of Basic Skills (ITBS) reading battery prior to the intervention and post-intervention. The results of the study indicated that students who received the oral reading fluency training significantly outperformed the students in the control group on comprehension measures. The authors concluded that their study provided evidence for the connection between fluency and comprehension. These findings were supported by a recent report of the National Reading Panel (2000), which conducted extensive meta-analyses.

Curriculum-based measurement. Many of the students who experience reading difficulties have problems in Chall's second stage of reading development. It is in this stage that students learn to read with accuracy and fluency. Because many students get "stuck" in this stage, researchers and teachers have seen the importance of addressing students' reading accuracy and fluency as a route to increasing their comprehension.

Based on much research that provided strong support for LaBerge and Samuels's theories regarding reading fluency and comprehension, researchers and instructors have concluded that, not only is it necessary to provide specific fluency instruction, it is necessary to assess the students' progress using Curriculum-based measurement (CBM). Not only is CBM a quick, efficient, and informative way to assess students' progress and tailor their instruction, it provides an easy way to deliver clear feedback to both students and teachers. Madelaine and Wheldall (1999) cite studies by Deno and Fuchs and Marston and Magnusson to describe the many benefits of CBM. Among the advantages

are increased content validity, data that is clear and easy to communicate to teachers and parents, responses from students that require production rather than simply selecting answers, more effective monitoring due to the frequency of measurement, cost- and time- effective measurement, and a way to measure both accuracy and fluency. In addition to these advantages, researchers point out that CBM is sensitive to even small changes in a student's progress. Unlike most standardized measures in which students' progress is judged in yearly or monthly increments, several studies on reading CBM report that gains in students' reading fluency can be seen weekly, with students gaining an average of 1-3 words per week, depending on their reading level (Deno, Fuchs, Marston, & Shin, 2001).

In addition to test sensitivity, the reliability and validity of CBM have also been well documented (Deno, Fuchs, Marston, & Shin, 2001). Madelaine and Wheldall provide a description of several studies to demonstrate the criterion validity of CBM. These studies showed stronger correlations between reading aloud measures (i.e., oral reading fluency) and standardized reading tests than other forms of reading measures (e.g., cloze). In addition, studies have shown strong correlations between read aloud measures and standardized reading comprehension tests. Kranzler, Brownell, and Miller (1998) attempted to examine the construct validity of CBM in reading by testing a hypothesis that surmised that certain children read more fluently than other children simply because they possess greater general intelligence, which enables them to process information more efficiently. This rivals the hypothesis that more fluent readers have better developed automatic decoding skills. Although the researchers included several

measures of general cognitive ability as well as processing speed, CBM oral reading fluency significantly predicted reading comprehension. The researchers concluded that this result is supportive of the construct validity of CBM for assessing oral reading fluency. Reliability can be demonstrated by correlating students' scores on similar passages. In addition, Marston and Magnusson (1988) point out that the standard error of measurement decreases with repeated administrations of CBM, which in turn increases reliability.

Schunk and Rice (1989) cited numerous studies that showed the importance of setting specific performance goals for students' performance. Schunk and Rice studied the effects of setting specific performance goals on the self-efficacy and reading comprehension of a sample of 4th and 5th grade students. The results of the study showed that students who were given specific, short-term, and challenging goals achieved higher scores on comprehension measures and rated themselves higher on their personal beliefs in their ability to understand what they read. This finding supports one of the major components of CBM; students set goals with the help of the instructor and monitor their own progress toward meeting these goals.

Self-Efficacy

Self-efficacy was a term first coined by Albert Bandura. He was convinced that there was more to behavior change than a simple response-outcome paradigm. He defined self-efficacy as "beliefs in one's capabilities to organize and execute the course of action required to produce given attainments" (1997; p.3). Bandura (1977) used this description to make an important distinction between outcome expectancy (i.e., what

response produces what outcome) and efficacy (i.e., one's belief that they are capable of successfully executing some behavior to produce an outcome). Self-efficacy has been researched extensively, both by Bandura and other researchers. Topics of interest have included self-efficacy related to various academic subjects, athletic skills, and career trajectories. Bandura (1997) points out that self-efficacy is not a phenomenon that can be measured by asking yes or no questions; rather, self-efficacy is judged on a continuum. For example, it is not informative to simply ask someone whether they believe they are capable of performing athletic activities. It is quite likely that, although someone does not see himself as particularly athletic, he judges himself to have more ability in a particular sport or sports requiring a particular skill such as the ability to strategize. Bandura stated, "The issue is not whether one can do them occasionally but whether one has the efficacy to get oneself to do them regularly in the face of varied dissuading conditions" (1997, p.43). One of the important conclusions that Bandura makes in describing his theory of self-efficacy is that if people do not feel capable of performing a particular task, they are likely to either not attempt the task or not put much effort into the task if it is attempted.

Bandura's theory and conclusions are particularly important to the field of reading research. If students repeatedly face a great deal of frustration when reading, they may perceive themselves as incapable of reading well. When students reach the conclusion that they cannot read well (i.e., they have low self-efficacy with regard to reading), it is unlikely that simply rewarding them with tangible objects or praise will be

enough to motivate them to engage in reading and work hard at trying to improve. There must be some cognitive mediation that motivates them to do so.

Schunk (1984) studied a sample of 33 elementary school students ages 9-11. He was interested in seeing whether providing students with proximal goals as well as performance-contingent rewards (both of which had been shown in previous studies to contribute to the performance, motivation, and self-efficacy of students) would result in higher self-efficacy and performance ratings than either condition alone on a task involving mathematics division problems. Schunk had students rate their self-efficacy on a scale ranging from 10-100, with 10 indicating high uncertainty and 100 indicating their complete confidence. Schunk trained the students to judge their abilities on a continuous scale such as this by having children judge their ability to jump longer and longer distances. The results of the study supported Schunk's hypothesis: The combination of performance-contingent rewards and proximal goals resulted in higher self-efficacy and better performance on the division problems than either condition alone. This is important because it shows that, not only do teachers need to provide relatively immediate goals (e.g., increase your correct words read correctly per minute from 75 to 77 by next Friday), rewards need to be based on the students performing some action based on a set criterion such as writing an essay having certain qualities such as all words spelled correctly, completing a certain number of mathematics problems correctly, or reading a paragraph fluently rather rewarding a student simply for completing the task or for participating.

In addition to studying connections between self-efficacy and academic achievement, researchers have studied how self-efficacy may influence future career paths of students. Bandura, Barbaranelli, Vittorio Caprara, and Pastorelli (2001) conducted a prospective study of 272 children aged 11-15 years in Rome, Italy. They were interested in whether children's beliefs in their ability with regard to academics, social situations, and self-regulation influenced their choice of occupation. In addition, they studied parents' self-efficacy and aspirations to try to identify a connection with their children's perceived beliefs in their ability to perform certain careers. The researchers found that there was an indirect link between the socioeconomic status of the family. In addition, they discovered that it was the children's beliefs in their own abilities that was the biggest influence on their beliefs in their ability to perform particular occupations as well as their preferences of a career. They found that girls had more confidence in their ability to succeed in education or health-related occupations while boys judged themselves to be more efficacious in science and technology-related fields.

One of the major conclusions that can be reached from studying existing literature on self-efficacy, is that it is intertwined with students' academic achievement as well as decisions they may make regarding their futures. Therefore, it is important not only to intervene and implement interventions to improve the reading of students at-risk for failure, it is crucial to implement interventions that are rewarding and provide challenging, yet attainable goals in order to promote the self-efficacy of these students

toward reading. In this way, we may hope to maintain and extend any progress the student makes.

Future research in self-efficacy should examine ways of promoting self-efficacy in various ethnic groups toward reading. In addition, although there is a plethora of information on various intervention strategies for improving reading fluency, accuracy, and comprehension, evidence is lacking that these interventions specifically promote the students' beliefs in their ability to perform the given task. Most of the self-efficacy research focuses on children in typical public school settings, mostly in general education classes, and in limited parts of the United States or other countries (e.g., Canada, Italy). Although there are many studies pertaining to the self-efficacy of teachers, few research articles exist that study the self-efficacy of students in special education classrooms. Investigating ways to promote the self-efficacy of students in alternative school settings and/or special education classrooms is an important and worthwhile area to study, which is virtually uncharted in the area of reading.

Social Comparison and Reading Performance

Social comparison theory is a term coined by Leon Festinger (1954). Social comparison is similar to self-efficacy in that both may affect the efforts and interests of a particular student. In essence, social comparison theory describes one's tendency to compare one's own abilities or opinions with those of others. Festinger took this basic theory further, describing several complexities as well as research support for his theory.

Among the many facets of social comparison theory is the concept that people are less attracted to situations in which others differ significantly from them in either

opinions or abilities. It is easy to see how social comparison may impact students with reading difficulties. According to this theory, poor readers would be less attracted to situations in which most others in the class would perform better than them (i.e., reading tasks). To tie this together with the previous discussion on the link between academic and behavior problems, it is plausible that students who perform poorly on reading tasks would also be more attracted to other students who perform poorly on reading tasks.

Another intriguing hypothesis raised by Festinger was that people will rely more on the opinions or comparative abilities of others if there is not some other objective criterion present with which to judge themselves. Framing this hypothesis within the reading domain, it would stand to reason that students often compare their own ability to read with others because there is not an immediate, objective criterion with which to compare their ability. Therefore, if the student has a tangible criterion (e.g., “last week I read 78 words correct in one minute, my goal this week is to read 80 words correct in one minute”), he may be more likely to judge his ability against his own criterion rather than trying to decide if his performance is better or worse than a particular peer. This part of the theory may have important implications for CBM. Although the purpose of any reading program is to improve students’ performance to be closer to the average peer, CBM operate by constantly providing feedback to each student about his or her own goals and achievement. Perhaps by doing this, students will rely less on comparisons of themselves with others and more on personal achievement to make judgments about their ability to read.

A longitudinal study by Blatchford (1997) in inner city London schools showed that students lowered their self-assessments of mathematics and reading ability from age 7 to 16 years. In addition, students' ratings became more accurate with age as demonstrated by comparing their ratings to performances on reading and mathematics measures. This study also supported previous research suggesting that students compare themselves using lower ratings when relating to peers in their own class rather than peers of the same age in general. Expanding on this research with normal-achieving students, Bear, Minke, Griffin, and Deemer (1998) discussed research that has found that children with learning disabilities are fairly accurate in their judgments about their own academic difficulties. Although researchers have consistently observed the accuracy of student's self-judgments, they have also found that children with learning disabilities are well aware of the importance of academic competence. Bear, Minke, Griffin, and Deemer sought to extend previous research with learning disabled (in reading) and normally-achieving students by studying students' perceptions of teacher feedback, social comparison, and reading satisfaction. They found that social comparison was more important to the 6th graders in their study than it was to the 3rd graders. This supported previous research that suggests that social comparison becomes more influential as students age. Perhaps the most interesting finding in their study was that there were no differences between students with learning disabilities and normally-achieving peers with respect to their perceptions of teacher feedback and reading satisfaction. However, they did find that the older students (i.e., 6th graders) who had learning disabilities judged themselves more accurately with regards to social comparison. These researchers

argue that it is desirable for students with learning disabilities to respond positively to reading, even if they know that they are poor readers. This is because reading is particularly difficult for these students and motivation often suffers. The researchers also conclude that positive teacher feedback can mediate the impact of social comparison as it becomes increasingly more important for older students.

These studies all show that, regardless of reading difficulty, older students (i.e., middle school to high school age) can be quite accurate in their judgments of their ability to read. These students may also have positive regard toward reading even though they realize they are poor readers. Finally, social comparison becomes more important to older students, and their self-ratings are lower when comparing themselves to peers in their class rather than peers their same age in general. This is important because, as discussed previously, there are many factors that contribute to reading difficulties and “giving up” behaviors. If students with reading difficulties constantly compare their reading performance to similar-aged peers in their classes, they may conclude that their reading is not improving and give up trying to improve their reading skills.

Read Naturally

“Read Naturally”(RN) is a strategy designed to improve the reading fluency of students who are struggling (below the 50th percentile of reading fluency compared to same-age peers). The RN strategy combines key elements of fluency instruction supported by research on reading fluency. These elements include modeling, repeated reading, and progress monitoring. In addition, the strategy is highly motivating (Hasbrouck, Ihnot, & Rogers, 1999; Ihnot, 2001), a key factor in developing reading

automaticity (Samuels, 1986). Students are fully involved in their own instruction by choosing stories of interest to them and graphing their progress. Although the focus of the RN program is on fluency, comprehension practice is included as well.

Prior to implementing the RN program, a trained teacher determined each student's instructional level. This was determined by having the students read CBM passages from a placement packet included in the RN materials. The teacher estimated the level based on knowledge of the student. The packet included a placement table, which helped the teacher determine at which level students should receive instruction. In addition, the table included information to help the teacher determine the students' goal.

There are twelve simple steps in the RN program: 1. The students select a story from a packet at the reading level that is appropriate based on informal assessment. 2. Next, the students consider what the topic might be discussing and write a few words at the top of the page. 3. The students then time themselves for one minute, reading the story aloud. If there are words that the students do not recognize, they underline them as they read. 4. The students then graph how many words they read correctly. 5. After the students graph their words correct per minute (wcpm) on the "cold" reading, they read along with the audio tape that corresponds with the story they chose. The read along is repeated three times. 6. After three times through with the audio tape, the students practice reading the story without the tape, timing each reading, until the passage is read at the their predetermined goal rate, determined by the students with the help of the teacher. 7. The next step is to answer the questions at the end of the passage. 8. The students have to "pass" the story by reading aloud for the teacher, who times the students

and determines whether they reached their goals. 8. The students graph their wcpm on the teacher-timed reading. 9. In another element that emphasizes the importance of reading for understanding, the students write a retell of the story. This involves writing a specified number of ideas learned from the story or writing for a specified amount of time. 10. After the written retell, the students pick out a new story at the same level of difficulty. 11. The previous 10 steps are repeated until the students complete 12 stories in a level or complete their stay at the school. 12. If appropriate, the students and teacher study the students' graphs and determine whether they should continue at the current level or adjust the goal or material difficulty.

Current Study

Although thousands of students attend alternative education settings every year due to behavior problems at their home schools, very little research has been conducted addressing the effectiveness of these programs for improving students' academics. There currently is no mandated system of assessing students' reading ability and no required programs of remediation. On top of the lack of mandated programming, there is not a required system of evaluation to determine if the alternative placements are effective. Researchers have documented a link between behavior problems, reading, and self-efficacy.

Therefore, this study sought to implement an oral reading fluency program to address this need at a disciplinary alternative education setting. In addition to training teachers and employing an oral reading fluency program, the following research questions were addressed: (a) Does a short-term, intensive reading intervention

targeting oral reading fluency improve the reading accuracy and reading fluency skills of students temporarily assigned to an alternative school due to behavior problems?, (b)

Does the reading intervention have an impact on reading comprehension of the students?, (c) How does this intervention affect the students' self-efficacy with regards to reading?, and (d) How does the intervention affect the way students compare themselves to others with regards to reading?

CHAPTER III

METHOD

Subjects

Participants in this study were 26 students in an alternative school in a mid-size city in Central Texas. There were 19 males and 7 females, ranging in age from 12.0 and 16.8 in grades 6 through 9. The ethnic make-up of the sample was as follows: 2 Caucasian, 9 African American, and 15 Hispanic. Sixteen of the students participated in special education classes, ranging from content mastery classes to total inclusion in special education and 1 student was enrolled in an English as a Second Language class.

Students qualified for participation in the study based on assessments conducted upon intake to the alternative school setting once parental consent was obtained for the assessment. Students were selected for screening based on record review of the Texas Assessment of Academic Skills (TAAS) scores, grades, and other relevant information such as participation in special education or other related services for reading intervention. Students evidencing reading deficits, as determined by performance at or below the 40th percentile on standardized reading assessments described below, were included as subjects. The reading levels of the subjects ranged from a grade level of 0.8 to 6.5. The average reading level of these students was 3.0 ($SD = 1.5$). The participants were assigned to the alternative school for a period of 20 to 40 or more days, based on rulings made by the school district's disciplinary committee, as well as their behavior while at the school. The average length of stay was 33.27 days ($SD = 7.06$).

Design

The design was a series of 26 single-case AB studies. The independent variable was the instruction program, in this case, a variation of Read Naturally. The dependent variables for this study were oral reading fluency (ORF), oral reading accuracy (ORA), comprehension, and self-efficacy toward reading. Oral reading fluency was measured proximally using curriculum-based measurement. In addition, ORF and ORA were assessed using external measures, including the Test of Word Reading Efficiency (TOWRE), Gray Oral Reading Test, Third Edition (GORT-III), and the Word Attack subtest of the Woodcock Reading Mastery Tests- Revised (WRMT-R). Because comprehension and self-efficacy were hypothesized here to be outcomes of improving ORF and ORA, they were each assessed using distal measures. Comprehension was measured using the GORT-3 and the Passage Comprehension subtest of the WRMT-R. Self-efficacy toward reading was measured using two rating scales developed by the researcher. In addition to the proximal and distal measures, it was expected that frequent student absences from school would cause lower weekly CBM scores. Thus absences were measured as a potential confound. See Figure 1 for an illustration of the design.

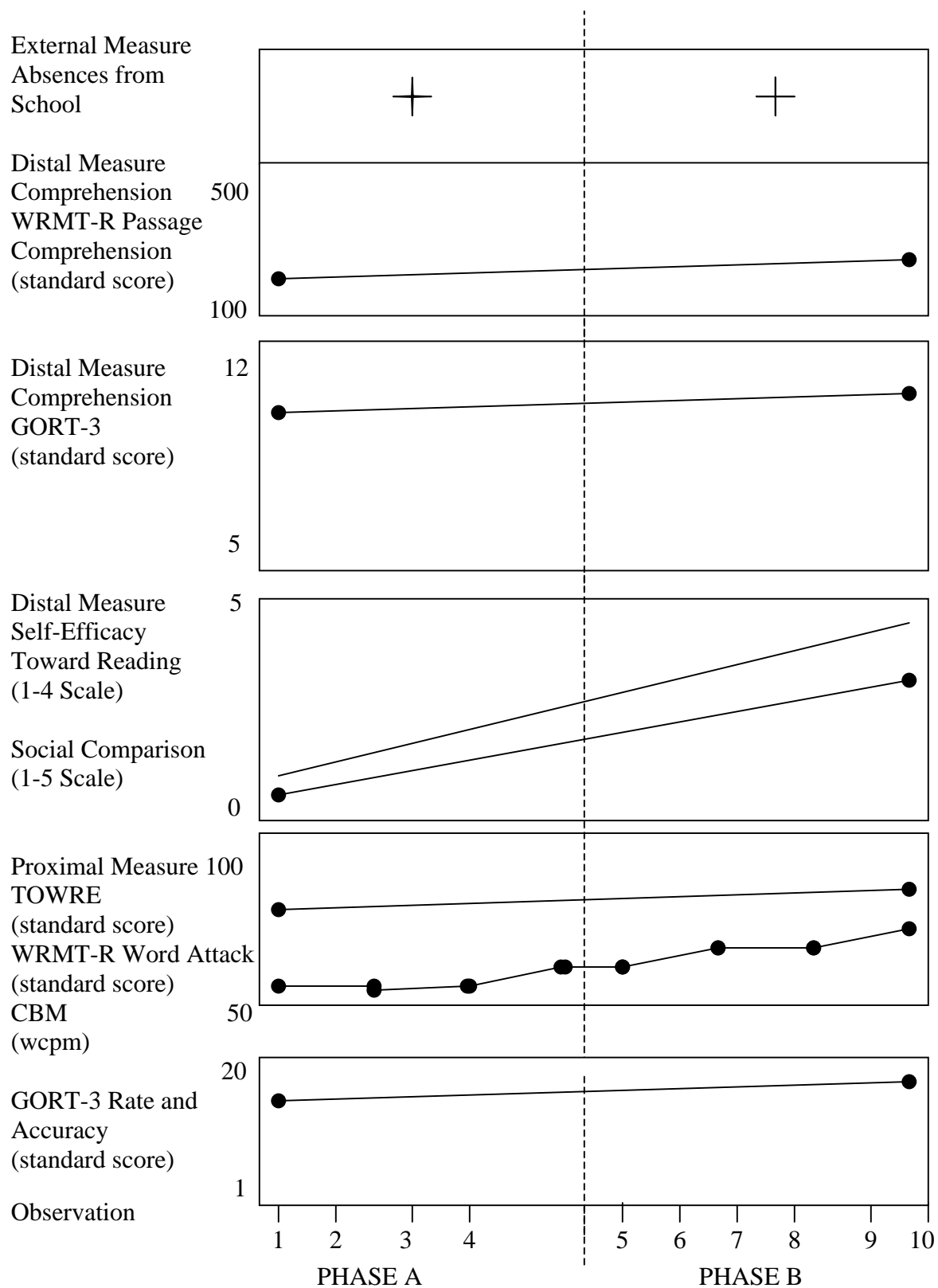


Figure 1. Schematic figure of study design.

Instruments

Pretest measures. All students entering the alternative school underwent a standard intake procedure. This procedure consists of a review of academic and discipline records as well as a formal interview by the school principal with each student and his or her parent(s). For this study, during the intake interview, the parent(s) of each student evidencing reading problems as determined by record review were asked for their child's participation in this study. Once signed consent and assent were obtained from the parent and student, additional assessments were conducted by trained graduate students.

Several reading assessments were administered. Although the main focus of the study was to measure improvement of reading accuracy and fluency, it was necessary to measure reading comprehension as well. Although the ultimate goal of reading is to understand what one has read, students who have weak decoding skills expend so much effort trying to guess and figure out what a word might be (and often guess incorrectly) that they do not understand what they have just read (Sanders, 2001). Therefore, improvement of decoding skills and reading rate should lead to improvement of comprehension skills as well.

Students were assessed using the Test of Word Reading Efficiency (TOWRE), a test that measures sight word and phonemic coding efficiency (Torgeson, Wagner, & Rashotte, 1999). The TOWRE provides normative information, including standard scores and percentiles, as well as psychometric data regarding sufficient reliability and validity.

The Woodcock Reading Mastery Tests-Revised (WRMT-R) is a test widely used to assess reading achievement and has sound psychometric properties (Woodcock, 1997). Two subtests of the WRMT-R were used. Word Attack was used to supplement information gleaned from the TOWRE. Although using one subtest of a larger test battery results in decreased reliability, reliability studies of the Word Attack subtest have resulted in correlations of .81 or higher. Because the TOWRE does not measure comprehension skills, students were assessed with the Passage Comprehension subtest of the WRMT-R. Reliability studies of the Passage Comprehension subtest resulted in correlations of .68 or higher.

In addition to the TOWRE and WRMT-R subtests, students' reading level was assessed using Gray Oral Reading Test, Third Edition (GORT-III). The GORT-III consists of two parallel forms with 14 passages in each. Passages are in developmental increments from 6-18 years of age and each passage includes 5 comprehension questions (Wiederholt & Bryant, 1992). A fluency score is obtained by taking into account the student's reading rate and accuracy. The GORT-III includes substantial reliability and validity data. Scores are reported as standard scores, percentiles, and age and grade equivalents. The students were placed in their instructional reading level (i.e., the level at which the student reads with some assistance) based on reading fluency and comprehension scores on the GORT-III as well as assessments conducted by the teacher using materials from the RN program packets.

Participants' self-efficacy was assessed individually using rating scales created by the principle investigator of this study (see Appendix A). The development of the

rating scales was based, in part, on previous research by Schunk (e.g., 1981), and Bandura (e.g., 1977). The Self-Rating of Reading scale was designed to assess students' self-efficacy with regard to passages the student may encounter in a typical social studies textbook at their grade-level. Students rated themselves in relation to same-age peers, answering questions about how sure they were that they could read a passage in about the same amount of time, make few mistakes while reading, and answer the same amount of questions correctly. After being given anchors and typical performance of peers, students judged how sure they were of their ability on a four-point scale with the following descriptors: Not Sure, Maybe, Pretty Sure, Real Sure. Students were then asked to think about their reading ability when they return to their regular schools and rate how sure they were that they could perform the following tasks: read a textbook out loud to their class without making mistakes, read a textbook chapter in front of their class in the same amount of time as other students, and read and understand their classroom textbook. Students rated how sure they were about performing these tasks on the same four-point scale described previously.

As a social comparison measure related to reading ability, the students then rated their ability to read compared to same age peers on a scale from 1-5, with anchors given at values 1 (students who cannot read the page at all), 3 (most students your age), and 5 (the best reader your age).

Weekly monitoring. Reading progress was monitored informally each week using Curriculum-based Measurement (CBM). CBM involves students reading a passage for one minute at their instructional level and graphing the results in order to monitor their

progress. It has been shown that students who set attainable goals and receive feedback on their progress toward meeting those goals have a higher self-efficacy (Schunk, 1984). Because students graphed their progress as a part of the intervention, they did not graph the weekly CBM measures.

Procedures for administering CBM were adapted from Shinn (1989). The examiner brought two copies of each passage at the students' goal level (usually one half to one grade level above the students' instructional level) to be read at the testing session. This enabled the examiner to read along and mark errors while the student read. The examiner said to each student, "When I say 'start', begin reading aloud at the top of this page. Read across the page. Try to read each word. If you come to a word you don't know, I'll tell it to you. Be sure to do your best reading. Are there any questions?" The examiner then said, "start". The examiner started his or her stopwatch as soon as the student began reading. As the student read, the examiner marked any errors. Mistakes marked as errors included omitting a word or letter, mispronouncing a word, and words prompted by the examiner. If a student had difficulty pronouncing a word and did not say the word or paused for 3 seconds, the examiner said the word for the student and counted it as an error. Mistakes not counting as errors were repetitions, self-corrections, and adding a word. A fluency score was obtained by subtracting the number of words read incorrectly from the total number of words read in one minute. An accuracy score was obtained by dividing the number of words read correctly by the total number of words read, then multiplying by 100 to get a percent.

Prior to implementation of RN, baseline data was collected. Baseline data was collected using equivalent passages at the students' goal level. Baseline data was collected every other day for 1 week and 1 day, resulting in 4 datapoints. To decrease the variability of scores on the one-minute timings, students completed 2 one-minute timings at each testing session. Their scores were averaged to obtain a more reliable score.

Previous research on CBM indicates that a student in grades 1-6 should be expected to improve by approximately 1-3 more words correct per week, with those reading at a lower grade level making bigger gains (Fuchs, Fuchs, Hamlett, Walz, et al., 1993).

Posttest measures. The TOWRE, WRMT-R subtests, GORT-III, and Self-Rating of Reading measures were re-administered post-treatment, using alternative forms in order to reduce testing effects.

Intervention

Students were instructed using a variation of the Read Naturally program, an empirically based reading program shown to increase reading fluency (Hasbrouck, Ihnot, & Rogers, 1999). Instruction was individualized, and involved each student (a) reading a passage along with an audio tape, (b) reading that same passage independently several times, and (c) self-monitoring daily progress by graphing his or her results. Students received instruction using this intervention 4-5 days per week for up to 50 minutes daily. The intervention continued for the length of each student's attendance at the alternative campus, ranging from 20 to 40 days, although some students were involved for up to 50 days. Due to

the nature of single case time series designs, each participant served as his or her own control group; therefore, no control group was included.

Procedure

Informed, written consent from the parents and written assent from the students was obtained during the standard intake interview that each student underwent with the principal upon entry to the school. Parents were assured that their children would not be excluded from the reading program due to lack of participation in the study and that they could withdraw from the study at any time. Participants were assessed by trained graduate students between October, 2001 and May, 2002. Pre-treatment assessments included the TOWRE, WRMT-R, GORT-III, and Self-Rating of Reading. The order of administration was random and parallel forms for all measures except the Self-Rating of Reading were counterbalanced to reduce testing effects.

Using results obtained from the TOWRE, WRMT-R, and GORT-III, as well as placement testing by the teacher using RN, students were placed in their appropriate reading levels. Baseline information was gathered by administering a CBM in each student's goal level every other day for one week and one day, resulting in 4 baseline datapoints. During the baseline period, students were trained on how to use the Read Naturally program and chart their progress. Because the training consisted of procedural tasks such as what was expected of them and where to find materials, it was believed that the training did not affect baseline data. During the baseline period, the students read from books or magazines independently during their reading class period or played educational computer games. Although CBM data was obtained weekly, the TOWRE,

WRMT-R, GORT-III, and Self-Rating of Reading were re-administered on the 30th day of intervention and on each student's final day at the school if they remained longer than 30 days.

Generalizability

It was intended that results from this study would be generalizable to students with similar backgrounds and difficulties (e.g., those with recurrent behavior problems due in part to reading difficulties). Because the study was a series of replicated single case studies, the subjects were clearly described, and the intervention was relatively easy to implement with training, the results should be generalizable to similar populations. Although Read Naturally has been shown to be effective with other students, this study may lend credibility to the effectiveness of using the program with students in short-term alternative school setting.

One limitation of this study was that it did not allow us to judge whether the students' skills transferred to learning in their home schools upon returning. In addition, although it is relatively easy to detect growth with regard to oral reading fluency and accuracy, comprehension improvement is more difficult to detect in the short-term. Therefore, it is possible that a student's fluency may have improved, but comprehension did not. Another limitation was that, although the teachers have been formally trained in the RN program, three different teachers delivered instruction to their assigned students. As is the case in many situations, some of the teachers were more adept at implementing and monitoring the intervention than others. Although checks were made by the

researcher to ensure the fidelity of instruction and extra support was provided when needed, it is possible that some students received more intensive instruction than others.

Statistical Analysis

Does a short-term, intensive reading intervention targeting oral reading fluency improve the reading fluency and accuracy skills of students temporarily assigned to an alternative school due to behavior problems? This question has two subparts. The first subpart pertains to measurements of reading fluency pre- and post-intervention. The TOWRE provides standard scores for Sight Word Efficiency, Phonemic Decoding Efficiency, and Total Word Reading Efficiency based on age. The authors of the GORT-III named their scale measuring fluency the Passage Score scale. Repeated measures (within-subjects) Analyses of Variance (ANOVA) was conducted on each of these scales to determine whether there was a statistically significant improvement on these five scales after the Read Naturally program is implemented for each subject. Because each scale is a result of converting raw scores to get standard scores, the data will be considered continuous.

The second subpart pertains to time series measures taken weekly through CBM. Progress monitoring data was analyzed using appropriate models, including mean level shifts between phase, general regression, and semi-partial correlation models described by Allison and Gorman (1993).

Simple Mean Difference. First, the Simple Mean Difference model was utilized for each of the 26 students. This model shows whether there was a change in mean level between the baseline and intervention phases for each subject, interpreted as an

improvement in oral reading fluency. The formula for the Simple Mean Difference model is:

$$Y_t = b_0 + b_1X_t + e_t$$

where Y_t is the dependent variable at time t , X_t is a dummy variable labeled zero for baseline and 1.0 for intervention phase, b_0 is the baseline mean, b_1 is the difference between the intervention and baseline means, and e_t is error (Gorman & Allison, 1997).

The Simple Mean Difference model is the most commonly used analysis in single case research (Parker & Brossart, 2003) and is useful when there are short phases. This is because trend is not considered; thus, data can be somewhat less stable. One of the disadvantages of this technique is that it lacks power, or sensitivity, to detect meaningful changes when there is an improvement in trend, meaning that improvement cannot be attributed to program effectiveness (Gorman & Allison, 1997). Therefore, to supplement this analysis, the Mean + Trend Difference (M + T) model was employed.

Mean + Trend Difference. The formula for the M + T Difference model is:

$$Y_t = b_0 + b_1t + b_2x_t + b_3x_tt + e_t$$

where b_0 is the baseline mean, b_1 is the trend for the entire series, b_2 is the difference in mean levels between phases, b_3 is the treatment interaction of the trendlines varying between phases, x_t is the dummy coding of zero for baseline and 1.0 for intervention, and e_t is error (Gorman & Allison, 1997).

This analysis takes into consideration mean and trend differences and enables the researcher to detect whether there was an improvement in a student's scores and whether this improvement may be attributed to the intervention. This is done by first determining

whether there was an improvement based on the effect size obtained using the Simple Mean Difference model. The $M + T$ model is then computed. If there is a large difference between the effect size obtained from the Simple Mean Difference model and the $M + T$ model, then likely the effect is due to the intervention.

Center Mean + Trend. Although the Simple Mean Difference and $M + T$ models are the most commonly used in single case research, statistical techniques have been devised which utilize semi-partialling or full partialling of regression lines to account for data trend are presumed to be more valid. Among these are the Allison Mean + Trend (Allison $M + T$; Allison & Gorman, 1993) technique and the Center Mean + Trend (Center $M + T$; Center, et al., 1985-1986) technique. The Allison $M + T$ technique is considered more valid due to semi-partialling out of only the baseline trend; however, this technique requires very stable baselines. With a short baseline, as is the case in this study, it is very difficult to establish a stable trend. One disadvantage of the Center $M + T$ technique is that it may actually underestimate treatment effectiveness because it may “over-partial” trend. For instance, if a subject’s trend line tends to slope upward, it is likely that it would continue upward through the intervention phase. The Allison and Gorman technique partials out this effect, leaving only the effect accounted for by the intervention. If the baseline data are not sufficiently stable to use this technique, other techniques such as one developed by Center, et al. (Faith, Allison, & Gorman, 1996) in which level, trend, and slope effects can be computed.

To determine whether the baseline phase was stable enough to utilize the Allison $M + T$ technique, ANOVA was conducted for each student to calculate the slope and

reliability for Phase A. Significance was determined to be $p < 0.10$. Out of the 26 students assessed, only 2 of their baseline slopes were determined to be stable.

Therefore, the Center M + T technique was used instead. The Center M + T technique is similar to the Allison M + T technique except overall data trend is partialled, rather than only the trend from Phase A. In addition, this technique can be used with short datasets. A major advantage of using this technique is that it tends to significantly reduce the problem of autocorrelation because mean and trend are both taken into consideration (Parker & Brossart, 2003). Autocorrelation basically is the amount of dependence one time sampling measure has on the previous measure. This can be problematic by inflating the effect sizes. For a more complete discussion on autocorrelation see Busk and Marascuilo (1992) or Gorman and Allison (1997).

Power analyses were conducted to determine expected effect size. For these analyses, the alpha level was set at $\alpha = .05$. This level was determined to be sufficient because it is unlikely that the intervention would be harmful to the students if it were deemed effective when in fact it is not. In fact, if the intervention is determined to show effectiveness, it could be very important in light of the seriousness of the issues facing some students at the school. However, it is still necessary to be somewhat conservative due to the cost entailed in materials, time, and training. Beta was set at $\beta = .20$, a number considered by Cohen (1988) to be acceptable. In order to determine the appropriate number of participants required to obtain a power of .80, a curve was calculated entering $N = 1$ to $N = 40$ by fives (see Figure 2). The analysis was conducted using two predictors because the Center Mean + Trend model includes two predictors

Multiple Regression Power Analysis

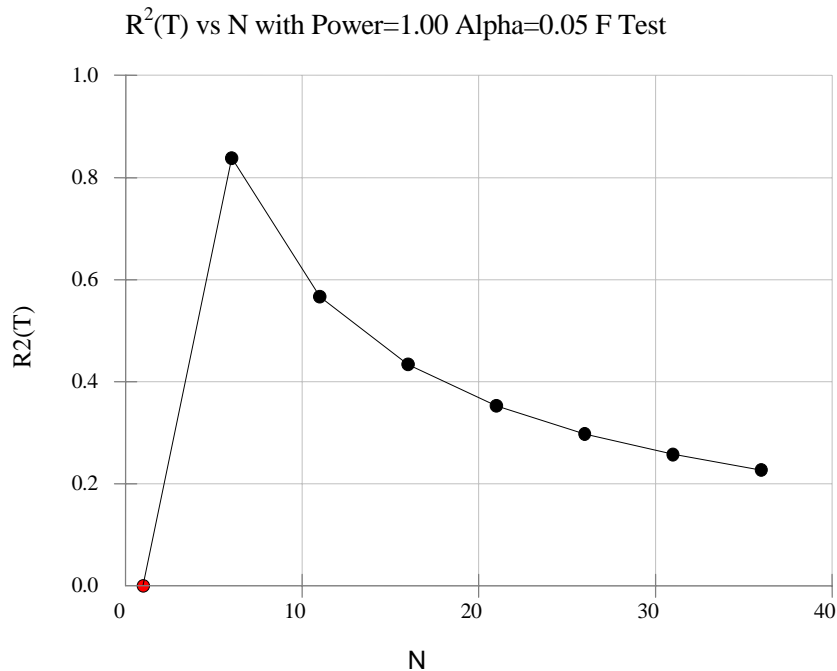


Figure 2. Power curve using multiple regression.

(trend and mean). When considering the above parameters, a sample of $N = 26$ enables reliable detection of effect size $R^2 = .15$.

Does the reading intervention have an impact on reading comprehension of the students? In addition to providing measures of reading fluency, the GORT-III provides standard scores for a reading comprehension scale. Because the GORT-III is essentially an informal reading inventory, the focus of the assessment is more on fluency than comprehension. Therefore, comprehension questions are limited as is the information obtained. Standard scores on the GORT-III and WRMT-R Passage Comprehension subtest

will be analyzed by using repeated measures ANOVA to determine whether there is statistically significant improvement on each of these two scales.

How does this intervention affect the students' self-efficacy with regards to reading?

The self-efficacy measure will be administered prior to and after intervention. Scores from the Self-Rating of Reading scales will be analyzed using repeated measures ANOVA to determine whether students rate themselves higher on self-efficacy toward reading after participating in Read Naturally.

How does the intervention affect the way students compare themselves to others with regards to reading? As with the self-efficacy measure, the social comparison measure (how students compare themselves to peers considering their reading ability) will be analyzed using ANOVA to determine whether students rate themselves closer in ability to higher-ability peers after intervention.

CHAPTER IV

RESULTS

Students were tested at an alternative education placement setting using measures of oral reading fluency, accuracy, and comprehension. The purpose of the study was to extend previous research on oral reading fluency interventions to a transient population of students placed in the alternative school due to behavior problems, as well as explore the link between reading and self-efficacy. This study was designed to answer the following research questions: (a) Does a short-term, intensive reading intervention targeting oral reading fluency improve the reading accuracy and reading fluency skills of students temporarily assigned to an alternative school due to behavior problems? (b) Does the reading intervention have an impact on reading comprehension of the students? (c) How does this intervention affect the students' self-efficacy with regards to reading? and (d) How does the intervention affect the way students compare themselves to others with regards to reading?

Fidelity checks were made by visiting the school at least one time per week to ensure proper implementation of the intervention. In addition, teachers, principal, and examiners were contacted at least one time per week in order to increase the chance that all aspects of the study were carried out as accurately as possible.

As stated previously, participants in this study were 26 students in an alternative school in a small city in Central Texas. There were 19 males and 7 females, ranging in age from 12.0 to 16.8 in grades 6 through 9. The ethnic make-up of the sample was as

follows: 2 Caucasian, 9 African American, and 15 Hispanic. Sixteen of the students participated in special education classes, ranging from content mastery classes to total inclusion in special education. Information on specific disabilities was not obtained. Students stayed at the school for an average of 33.27 days. The mean reading level at the beginning of the study was 3.02. The average grade level was 7; this indicates that, on average, students were about 4 grades below expected reading level. See Table 1 for means, standard deviations, and ranges.

Table 1.

Means and standard deviations of age, grade, reading level, and days stayed.

	Minimum	Maximum	Mean	SD
Age	12.00	16.80	14.00	1.19
Grade	6.00	9.00	7.19	1.13
Reading Level	0.80	6.50	3.02	1.53
Days Stayed	20.00	50.00	33.27	7.06

Descriptive Statistics

In order to summarize the distribution of the scores for each of the pretest measures, violin plots were included for each measure (see Figures 3-11). A violin plot is similar to a box plot, except a density trace is added to show the shape of each distribution (Hintze & Nelson, 1998). The density trace can be adjusted, and in this case was set at 5% in order to replicate the shape of the distribution of raw data almost identically. The median is displayed as a dark circle. The heavy line represents the interquartile range (25th to 75th percentile). It is clear from studying the plots that there was a considerable amount of

variability in the students' scores on each of the pre-test measures. For ease of comparison, each of the measures with means of 100 and standard deviations of 15 were plotted on one graph (see Figure 12) and the subtests of the GORT-III were plotted on another graph (see Figure 13) as their means and standard deviations were different ($X = 10$, $SD = 3$).

Students' median scores were below average across the measures, with the highest median score on the WRMT-R Word Attack subtest and the lowest median score on the GORT-III ORQ. Note that on assessments with a time component (i.e., fluency), students obtained lower median scores.

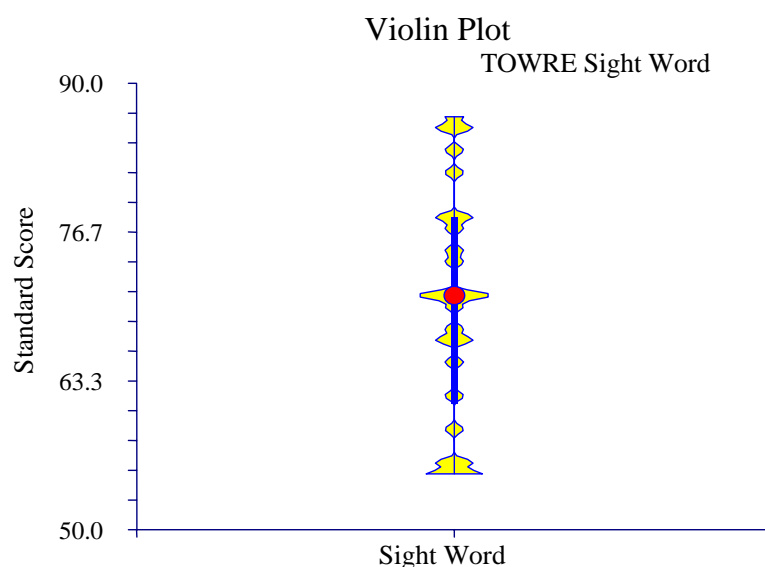


Figure 3. Violin plot of Test of Word Reading Efficiency Sight Word subtest (mean = 100, $SD = 15$). The quartile values were as follows: 10th = 55 , 25th = 61.25 , 50th = 71 , 75th = 78 , and 90th = 86. Three students obtained the lowest score on the distribution, while 4 students obtained the highest two scores (i.e., 86 and 87). All scores were below the subtest mean score of 100.

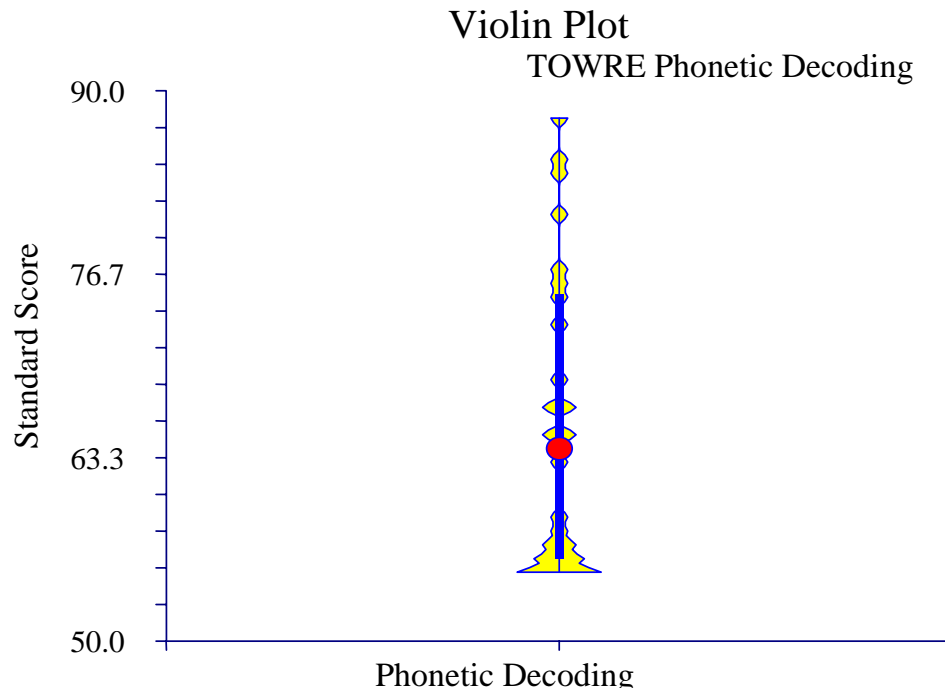


Figure 4. Violin plot of the Test of Word Reading Efficiency Phonetic Decoding subtest. This subtest has a mean of 100 and standard deviation of 15. The quartile values were as follows: 10th = 55 , 25th = 56 , 50th = 64 , 75th = 75.25 , and 90th = 84.3. Five students were grouped at the bottom of the distribution, while 1 student obtained the highest score (i.e., 88).

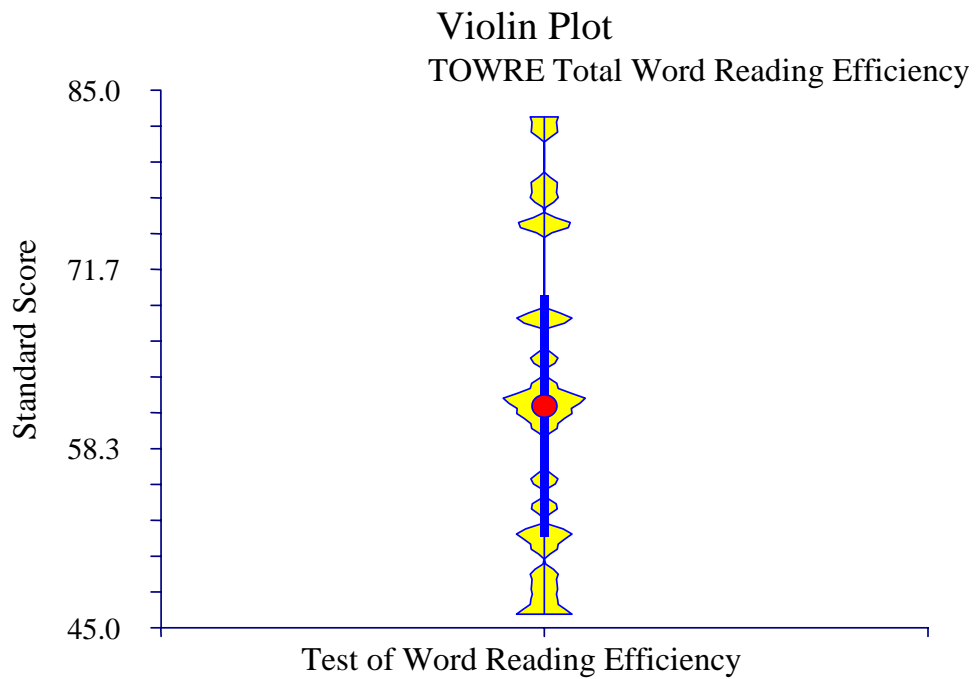


Figure 5. Violin plot of the Test of Word Reading Efficiency Total Word Reading Efficiency subtest. This score has a mean of 100 and a standard deviation of 15. The quartile values were as follows: 10th = 46.7, 25th = 51.75, 50th = 61.5, 75th = 69.75, and 90th = 79.2. Two students obtained the lowest score, and 2 students obtained the highest score (i.e., 83).

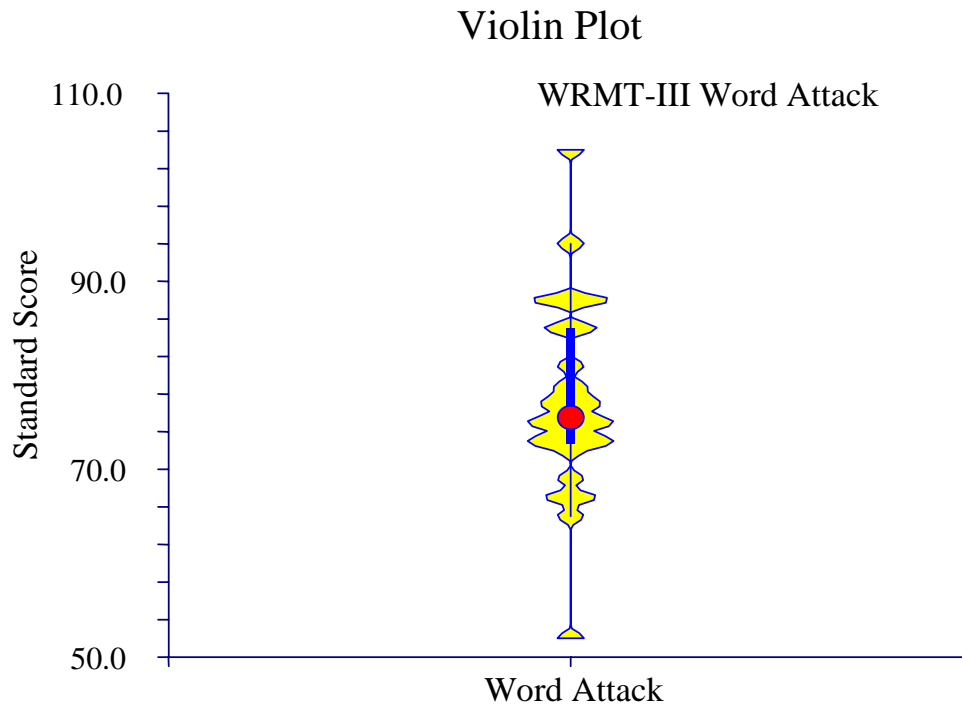


Figure 6. Violin plot of the Woodcock Reading Mastery Test-Revised Word Attack subtest. This subtest has a mean of 100 and standard deviation of 15. The quartile values were as follows: 10th = 66.4, 25th = 72.75, 50th = 75.5, 75th = 85, and 90th = 89.8. One student obtained the lowest score, and 1 student was at the top of the distribution (i.e., 104).

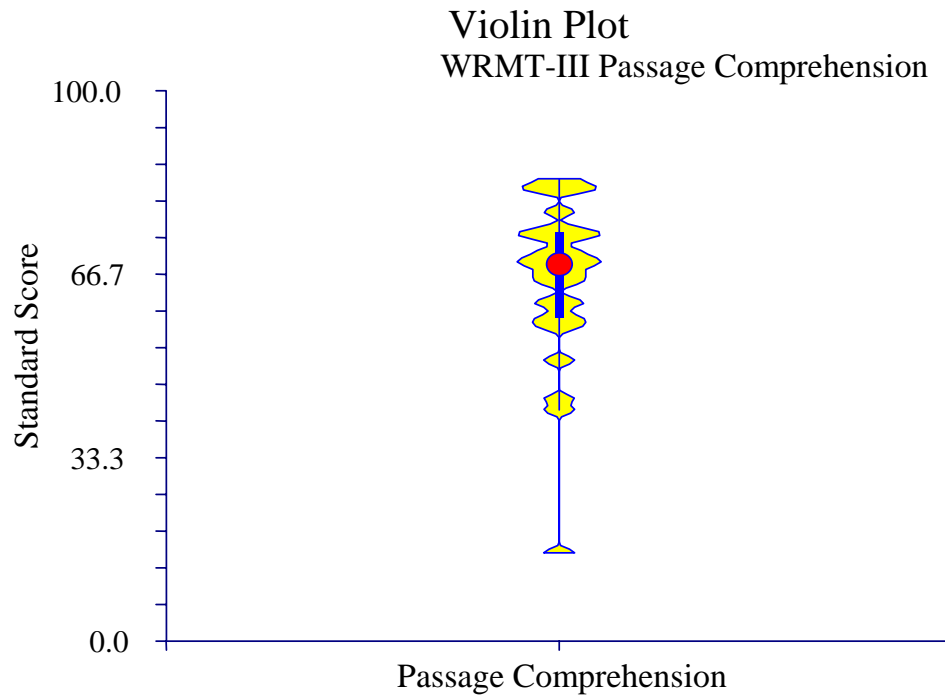


Figure 7. Violin plot of the Woodcock Reading Mastery Test-Revised Passage Comprehension subtest. This subtest has a mean of 100 and standard deviation of 15. The quartile values were as follows: 10th = 43.4, 25th = 58.75, 50th = 68.5, 75th = 74.25, and 90th = 82.3. One student obtained the lowest score on the distribution while 4 students were distributed at the high end of the distribution (i.e. between 82 and 84).

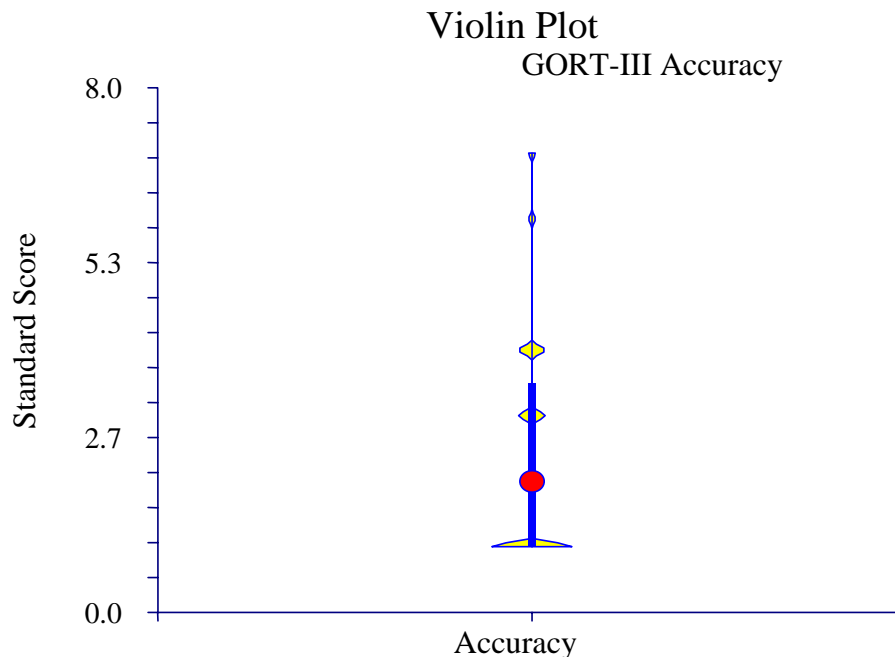


Figure 8. Violin plot of the Gray Oral Reading Tests Third Edition Accuracy subtest. This subtest has a mean of 10 and standard deviation of 3. Twelve students obtained the lowest score, while 1 student was at the top of the distribution with a score of 7. One score was missing due to the refusal of the student to complete the assessment.

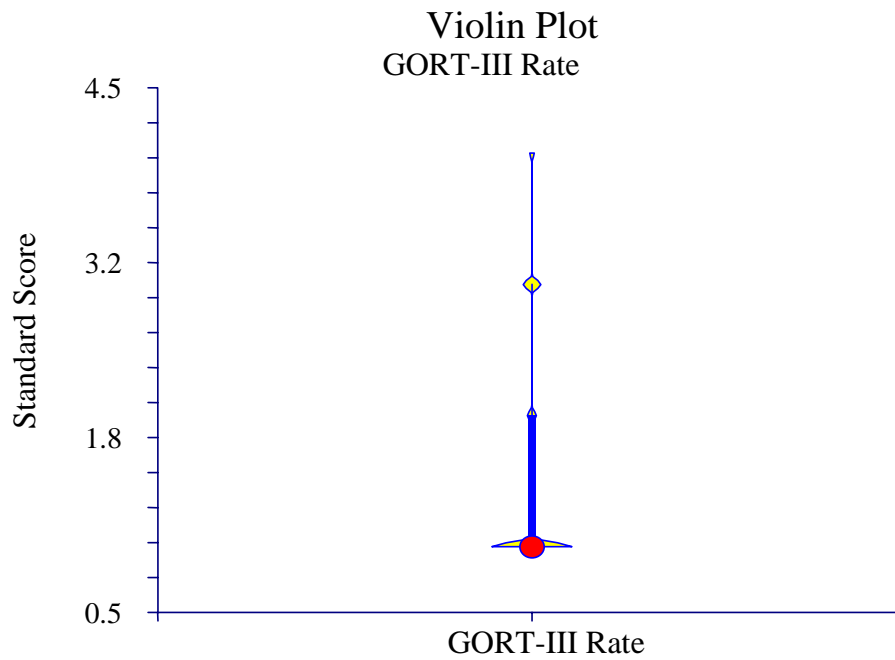


Figure 9. Violin plot of the Gray Oral Reading Test Third Edition Rate subtest. This subtest has a mean of 10 and standard deviation of 3. Eighteen students were grouped at the bottom of the distribution, while 1 student obtained the highest score of 7. One score is missing due to the refusal of the student to complete the assessment.

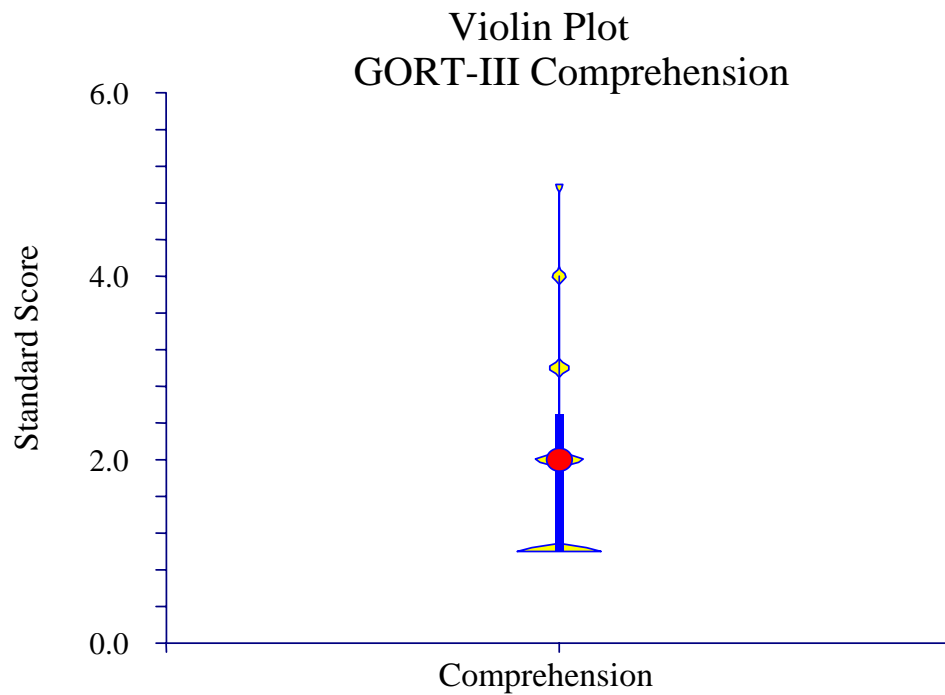


Figure 10. Violin plot of the Gray Oral Reading Test Third Edition Passage Comprehension subtest. This subtest has a mean of 10 and standard deviation of 3. Sixteen students were grouped at the bottom of the distribution with scores of 1 and 2, while 1 student obtained the highest score of 11. One score was missing due to the student's refusal to complete the assessment.

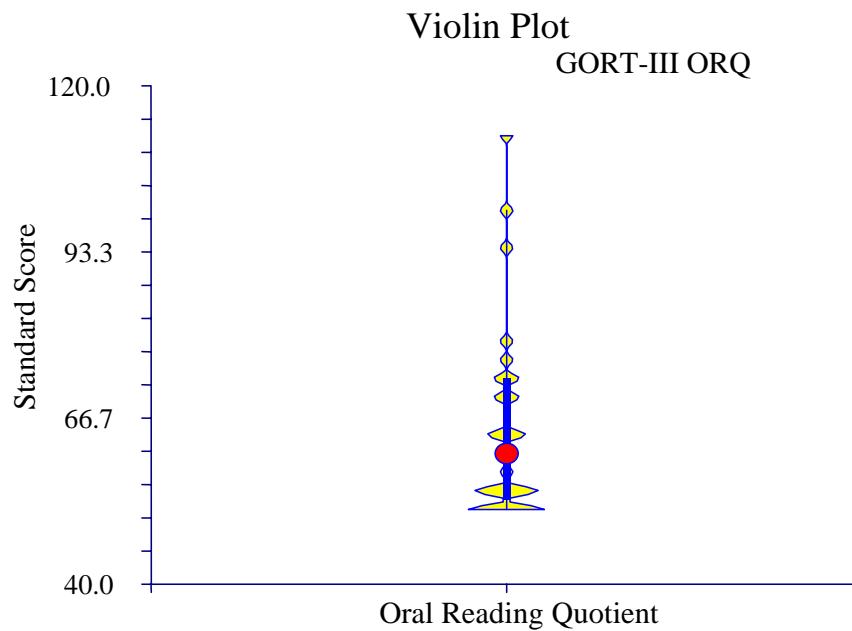


Figure 11. Violin plot of the Gray Oral Reading Quotient Third Edition Oral Reading Quotient. This score has a mean of 100 and standard deviation of 15. The quartile values were as follows: 10th = 52, 25th = 53.5, 50th = 61, 75th = 73, and 90th = 96.4. Six students were grouped at the bottom of the distribution, while 1 student obtained the highest score of 112. One score was missing due the student's refusal to complete the assessment.

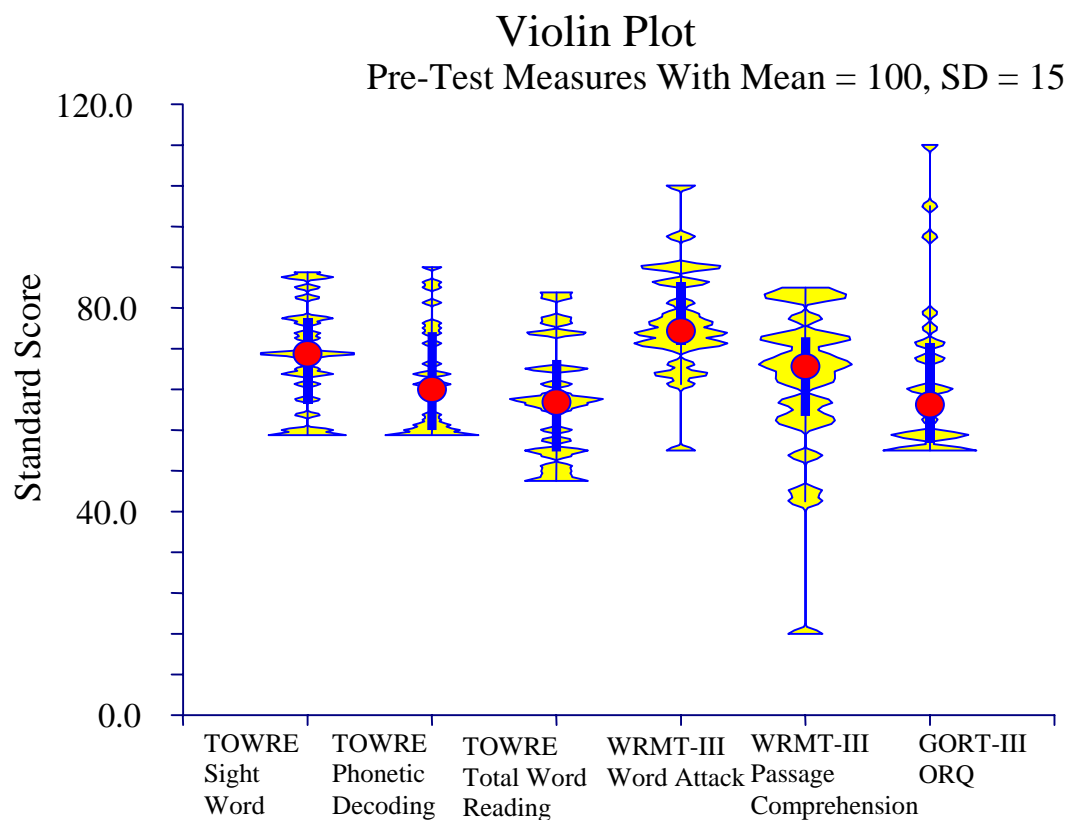


Figure 12. Violin plot of the pretest measures, excluding the individual subtests of the Gray Oral Reading Tests Third Edition. All measures included in this figure have means of 100 and standard deviations of 15. Note that students' median scores were lower on assessments that had a time element (i.e., TOWRE tests and GORT-III). The most variability in scores was seen on the two measures that included a comprehension component (i.e., WRMT-R Passage Comprehension and GORT-III ORQ). All median scores were well below average.

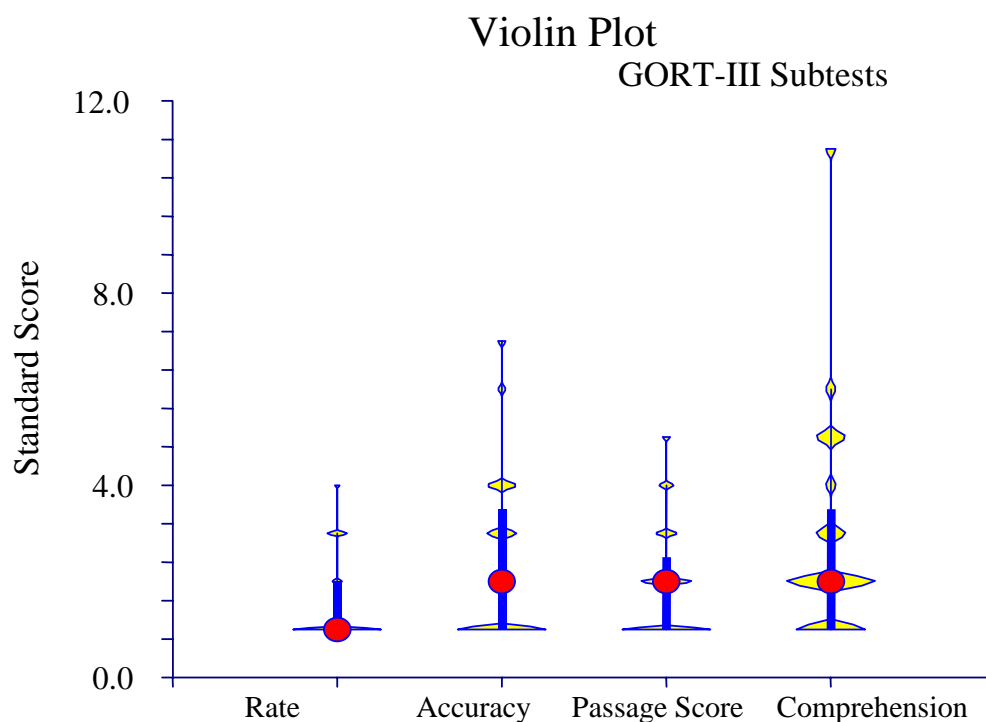


Figure 13 .Violin plot of the Gray Oral Reading Tests Third Edition subtests. These subtests each had a mean of 10 and standard deviation of 3. Note that the most variability was seen on the comprehension subtest. Similar to the pretest measures discussed previously, the lowest median score was on the rate subtest, which included a time element.

For the oral reading accuracy (CBM) measure, average baseline and intervention phase scores were computed for each of the 26 subjects by adding their scores in each phase (baseline and intervention) and dividing by the number of data points in each phase. The baseline scores were divided by 4 (with the exception of one student who had only 3 baseline datapoints) and the intervention phase datapoints varied from 1 to

10. These data were displayed in a violin plot (see Figure 14) in order to show the amount of variability present in each phase.

The lowest score in each phase was approximately 5 words correct per minute (wcpm), while the highest score increased from baseline to intervention phase by 12 wcpm. The average wcpm was approximately 61 in the baseline phase and 64 in the intervention phase. The wcpm in both phases were well below the 25th percentile compared to similar-aged peers (Hasbrouck & Tindal, 1991). Important to note are the large SDs. This is an indication of the amount of variability present in the scores. As discussed later, the amount of variability or “bounce” present was observable in the visual analysis as well.

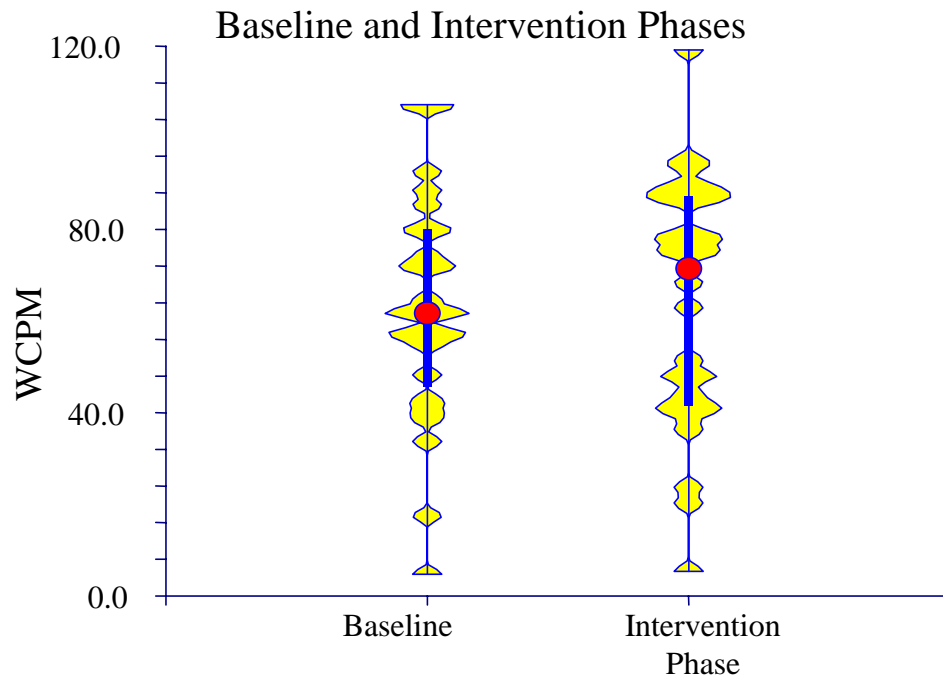


Figure 14. Violin plot of the baseline (Phase A) and intervention (Phase B) phases for average curriculum-based measurement scores for 26 students. Scores are reported in words correct per minute (WCPM). As expected, the median WCPM was higher in the intervention phase.

Each research question was addressed separately, with results following each question.

Does a short-term, intensive reading intervention targeting oral reading fluency improve the reading accuracy and reading fluency skills of students temporarily assigned to an alternative school due to behavior problems? To answer this question, two analyses were performed. The first was on the proximal measures, curriculum-based

measurement of oral reading fluency and accuracy, using visual and statistical single case analysis techniques. The second analysis was on the distal measures, TOWRE, WRMT-R, and GORT-III, using repeated measures (within-subjects) ANOVA Reading fluency skills.

Proximal Measure: Curriculum-based Measurement

The first analyses to be conducted were those pertaining to the students' oral reading fluency. The time-series data for each student were analyzed both visually and statistically.

Visual analysis. Parsonson and Baer (1992) described six main benefits of visual analysis. It is relatively fast and easy, does not require technology, and because graphs are so straightforward, there are not the theoretical debates that tend to surround statistical techniques. Visual analysis is particularly important as a supplement to statistical analysis because it is possible for a subject to demonstrate a large effect size, yet without examining the graph of this effect is impossible to determine whether the subject's performance improved or deteriorated. This is because effect sizes are not reported in negative numbers.

Each graph was judged based on 4 criteria well accepted in the field of single case research. These are bounce (variability), slope of the trendlines (regression), mean level, and difference in intercepts (the point where each slope hits the line of intervention). The researcher considered the criteria as a whole when judging whether there appeared to be improvement. One noticeable pattern was that each of the graphs had a considerable amount of bounce or variability, as noted above. This is not

surprising considering the relatively few data points obtained (and the short time frame). Examples of graphs showing low and high amounts of variability are displayed in Figure 15. These graphs are representative of the typical amount of bounce seen in the 26 graphs.

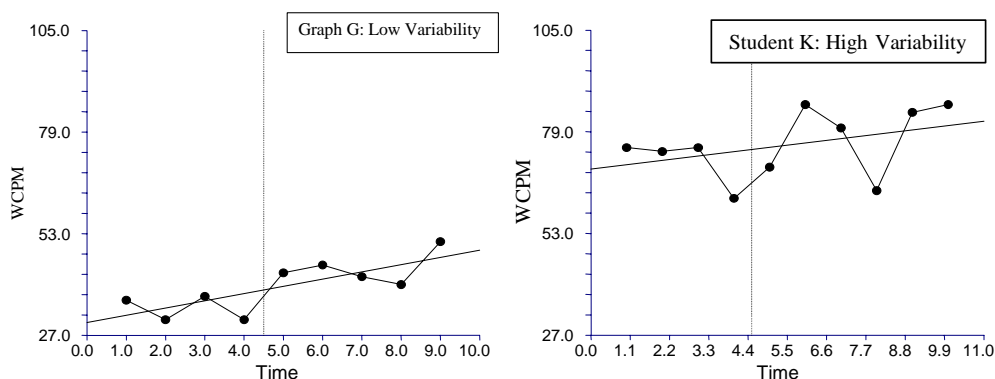


Figure 15. Fluency graphs depicting a low variability, or bounce (left), and high variability (right). Graphs were labeled for reference to the full set of graphs in Appendix B.

Because of the amount of bounce (21 out of 26 graphs appeared to have a great amount of variability), conclusions regarding treatment effectiveness and improvement in reading fluency can be only tentative in many graphs; therefore, caution was observed by not drawing strong conclusions based on the other three visual analysis techniques. This within-phase variability is taken into account during visual analysis by making judgments of improvement or intervention effectiveness more tentative. This variability is also considered in the statistical analyses in the calculation of effect sizes. Visual analysis was conducted by the researcher and corroborated by one other professional

experienced in visual analysis of time-series data. Taking into account the four visual analysis techniques discussed previously (i.e., shifts in mean, trend, and intercept), it appeared that 9 students showed improvement in mean level from baseline to intervention phase (see Figure 16 for an example). See graphs of the raw data for each student in Appendix B. Graphs depicting improvement are identified as following: C, D, E, G, H, J, K, N, and R

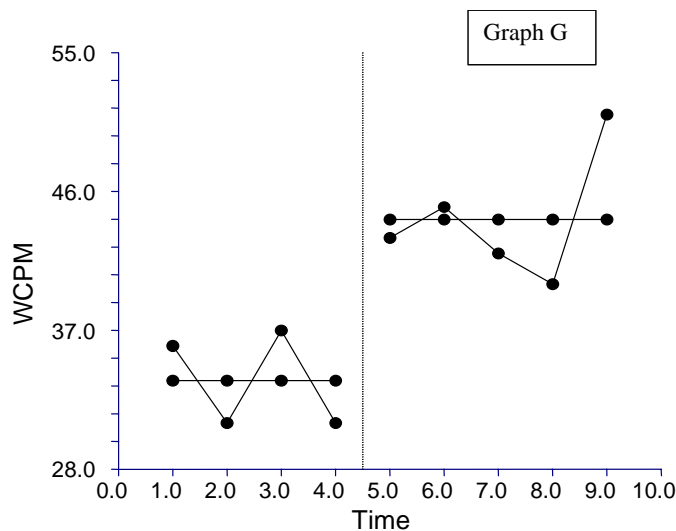


Figure 16. Graph of oral reading fluency illustrating improvement in mean levels between phases based on visual analysis. The mean level is indicated by a horizontal bar. The mean level of Phase B (i.e., intervention phase) is much higher than the mean level of Phase A (i.e., baseline).

The remaining 18 graphs appeared to indicate no improvement in mean level. In Figure 17, the graph on the left (Graph Y) shows no change in mean level, while the graph on the right (Graph L) depicts a negative mean level shift between phases.

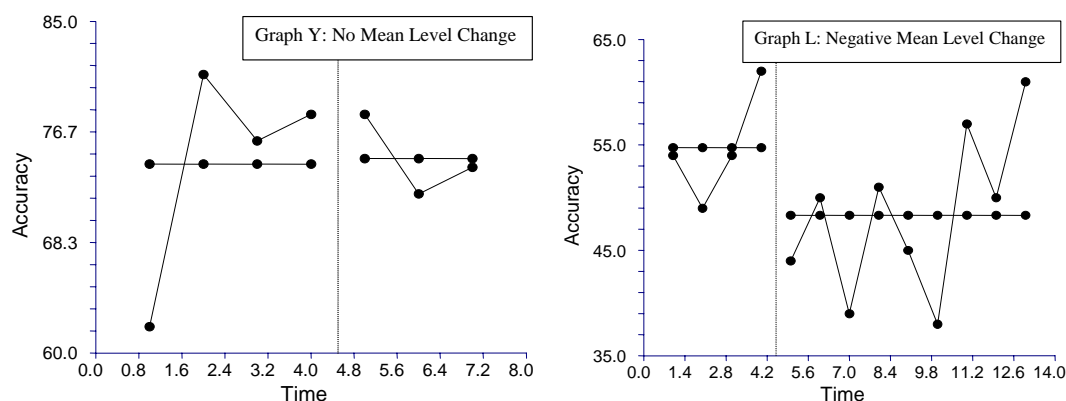


Figure 17. Graphs of oral reading fluency depicting no change in mean level between phases (left), and negative mean level changes (right).

Thus, based on visual analysis alone, roughly one third of the students showed improvement in their oral reading fluency scores from baseline to intervention. Slightly more than half of the students showed either no improvement or actually decreased performance in their scores. Studying changes in slope and intercept, 7 of the 9 graphs that show improvement in oral reading fluency show that this improvement was likely due to the intervention.

Statistical analysis. After visually inspecting the graphs of the oral reading fluency data, three statistical analyses were employed to enhance visual analyses. Statistical analyses yield effect sizes and confidence intervals, which allows for a more defensible interpretation of the results. These statistical techniques take into account data variability and the limited number of data points. In addition, autocorrelation will be studied to help examine the interpretability of the obtained effect sizes.

Simple Mean Difference and Mean + Trend Difference models. First, the Simple Mean Difference model was utilized for each of the 26 students. This model shows whether there was a change in mean level between the baseline and intervention phases for each subject, interpreted as improvement in oral reading fluency. Next, the Mean + Trend Difference (M + T) model was used. This analysis takes into consideration mean and trend differences and enables the researcher to detect whether there was an improvement in a student's scores and whether this improvement may be attributed to the intervention. Judgments of improvement and intervention effectiveness are preceded by first determining whether there was an improvement using the Simple Mean Difference model. The M + T model, which takes into account mean and trend differences simultaneously, was then computed.

Effect sizes and significance values for the Simple Mean and M + T Difference models are reported in Tables 2 and 3. Although significance values are often valued in group research, (Marascuilo & Busk, 1988), with $n = 1$, short time series make it difficult to reach statistical significance. Therefore, effect sizes tend to be more useful for interpreting the results in single case research, than simply noting statistical significance. As noted in Table 2, when considering the change in mean level between phases, only 8 of the 26 studies resulted in effect sizes in the low to moderate range (i.e., $R^2 = .30$ and above). Of these, the mean level of 3 students actually decreased in the intervention phase, leaving 5 students with improved mean levels.

When using the Simple Mean Difference model, autocorrelation can be problematic. In essence, autocorrelation occurs when residuals (i.e., error) in time series

data are correlated or related—in other words, not random. This can be problematic because a positive autocorrelation can lead to an inflated R^2 , while a negative autocorrelation can deflate the R^2 (Parker and Brossart, 2003). Autocorrelation is particularly problematic for techniques such as the Simple Mean Difference, because trend is not considered. Because analyses utilizing trend tend to minimize the autocorrelation problem, autocorrelation was examined for only the Simple Mean Difference analyses. Since 12 out of the 26 students had $r = .25$ and above (Scheffé, 1959), data was “cleansed” using the ARIMA (integrated autoregressive, moving average) Box-Jenkins technique then reanalyzed using the Simple Mean Difference to reduce autocorrelation. These results are included in Table 2. Notice that cleansing the data resulted in larger effect sizes for some students and decreased for others. Despite a slight increase in effect size for some students, there were still 8 out of 26 students with low to moderate effect sizes post-cleansing. Of these, 3 students showed a decrease in mean level in the intervention phase, leaving 5 students with increased mean levels post-intervention.

Table 2.

Effect sizes, pre- and post-cleansing of fluency for 26 students using Simple Mean

Difference model with 85% confidence intervals and standard errors.

Simple Mean R^2	Lower Limit	Upper Limit	SE	R^2 Post-Cleansing	Graph ID
0.001	0	0	-0.001	0	Y
0.008	0	0.197	0.063	0.001	X
0.010	0	0.237	0.075	0.023	I
0.015	0	0.247	0.075	0.004	P
0.023	0	0.286	0.083	0	B
0.025 ^a	0	0.324	0.095	0.077	V
0.026	0	0.346	0.102	0.405	F
0.080 ^a	0	0.433	0.0948	0.088	M
0.143 ^a	0	0.534	0.086	0.116	T
0.149 ^a	0	0.565	0.093	0.080	A
0.154	0	0.598	0.101	0.245	W
0.170	0	0.521	0.063	0.163	K
0.170 ^a	0	0.477	0.048	0.150	L
0.184	0	0.519	0.052	0.128	H
0.204	0	0.500	0.032	0.282	J
0.215	0	0.601	0.059	0.286	U
0.230	0	0.636	0.061	0.286	R
0.259	0	0.636	0.041	0.270	N
0.310	0.012	0.613	0.002	0.364	E
0.335 ^a	0	0.672	0.001	0.575	O
0.345	0.001	0.664	-0.009	0.362	D
0.360	0	0.772	0.018	0.273	Z
0.369	0.043	0.656	-0.014	0.378	C
0.411 ^a	0.036	0.707	-0.027	0.461	S
0.447 ^a	0.007	0.756	-0.045	0.595	Q
0.705	0.293	0.876	-0.084	0.766	G

Note: For ease of interpretation, the standard error (SE) is an average because the asymmetrical confidence interval is calculated with two SEs. This is a peculiarity of the F distribution. The graph IDs are included to permit comparison with the graphs in Appendix B.

^a denotes subject whose mean level decreased (deteriorated) from Phase A to Phase B

Table 3.

Effect sizes of fluency for 26 students using Mean + Trend Difference model with 85% confidence intervals and standard errors.

Mean Plus Trend R^2	Lower Limit	Upper Limit	SE	Graph ID
0.097 ^b	0	0.287	0.032	F
0.113	0	0.340	0.040	V
0.141	0	0.388	0.037	B
0.276	0	0.549	-0.001	P
0.295 ^b	0	0.583	-0.002	T
0.313	0	0.567	-0.020	E
0.317 ^b	0	0.603	-0.011	U
0.350	0	0.623	-0.027	O
0.352	0	0.617	-0.030	K
0.362	0	0.642	-0.028	Z
0.411	0	0.663	-0.055	S
0.435 ^b	0	0.705	-0.057	Y
0.437	0.024	0.662	-0.065	L
0.460	0.017	0.686	-0.075	C
0.487 ^b	0	0.746	-0.080	W
0.508	0	0.739	-0.096	M
0.524	0.122	0.720	-0.072	J
0.527	0.004	0.745	-0.106	D
0.546 ^b	0.062	0.751	-0.097	H
0.561	0	0.780	-0.119	Q
0.631 ^b	0.059	0.815	-0.135	X
0.692	0.064	0.854	-0.162	N
0.752	0.244	0.882	-0.131	G
0.814	0.181	0.919	-0.183	R
0.817 ^b	0	0.924	-0.247	I
0.859	0.31	0.940	-0.163	A

Note: For ease of interpretation, the standard error (SE) is an average because the asymmetrical confidence interval is calculated with two SEs. This is a peculiarity of the F distribution. The graph IDs are included to compare with the graphs in Appendix B.

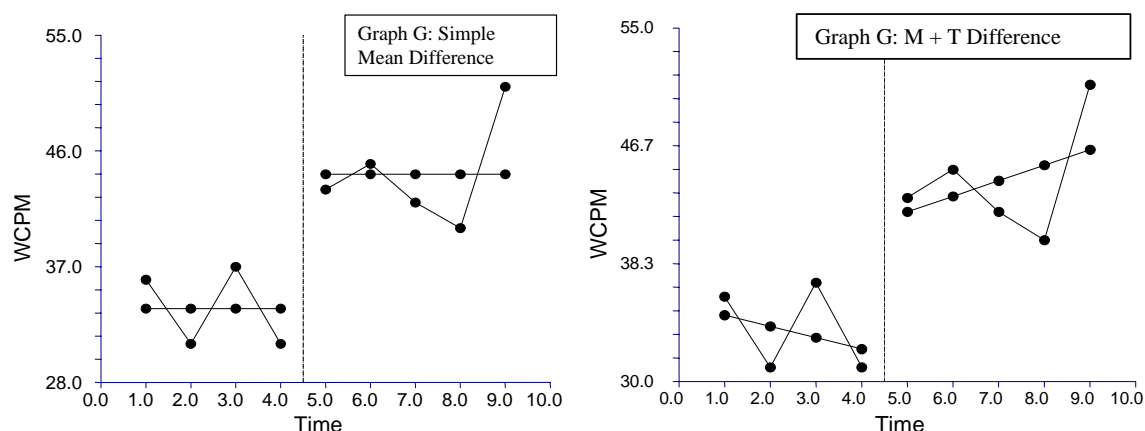
^b denotes subject whose Phase B trend was actually in the negative direction. Subjects with positive mean differences, but negative trends are anomalies. It is often recommended that M + T analysis is not computed in these cases as interpretation is not straightforward.

Although 8 students had low to moderate effect sizes (i.e., .30 to .80) when using the Simple Mean Difference model, 8 students actually had lower mean levels in the intervention phase, indicating that their performance deteriorated. In addition, it is evident from the confidence intervals as indicated by the upper and lower limits, that there is quite a bit of variability in most of the students' scores.

When taking into consideration trend differences (see Table 3), 18 out of 26 studies resulted in low to strong effect sizes (i.e., $R^2 = .30-.80$) and 3 students had large effect sizes (i.e., $R^2 = .80$ and above). Of these, 5 had positive increased trends from baseline to intervention phase, 11 had trends that decreased, and 5 had trends that did not change between phases. Of the 7 studies showing moderate to strong effect sizes (i.e., $R^2 = .50$ and above), using the M + T analysis, one study actually showed a decline in the student's performance, as indicated by a negative trendline. It is clear from the table that adding the element of trend in the computations produced higher effect sizes than considering only mean level. Of the 21 students who showed low to strong effect sizes, the performance of 6 students actually deteriorated. Although the effect sizes were increased by adding trend to the equation, there is still a considerable amount of variability as evidenced by the confidence intervals.

As mentioned previously, it is important to graph the analyses in order to compare the statistical results with visual analysis and properly interpret the results. Graphs of the CBM measure for one student obtained utilizing the Simple Mean and M + T Difference models are depicted in Figure 18. Note that the first graph shows simply that there appeared to be improvement in the student's scores between the baseline and

intervention phase. The second graph shows improvement in trend, and therefore, that the improvement was likely due to the intervention. See Appendix B for graphs of all 26 students' oral reading fluency scores using the Simple Mean and M + T Difference models.



Note: $R^2 = .705$

Note: $R^2 = .752$

Figure 18. Examples of fluency graphs using the Simple Mean Difference and Mean + Trend techniques for Student G.

Center Mean + Trend Difference. Although the Simple Mean Difference and M + T models are the most commonly used in single case research, presumably more valid statistical techniques that have been devised which utilize semi-partialling or full partialling of regression lines to account for data trend. Among these is the Center Mean + Trend (Center M + T; Center, et al., 1985-1986) technique. Results of the Center M + T analyses are reported in Table 4. Because the graphs for the Center M + T technique use detrended data, rather than raw data, visual interpretation is more difficult and less

straightforward. Therefore, for this technique, graphs are not included. As discussed previously, this technique partials out trend from the entire equation, not just the baseline. This is because in very short baseline phases, as was the case in this study, the baseline trend is not trustworthy (i.e., it is unstable). Because the M + T technique did not account for the short baseline trend, results from the Center M + T technique should be considered more trustworthy in this case. Although partialling trend from the entire series can result in an underestimation of the treatment effect, the high variability of scores and short datasets in this study require conservatism.

According to Parker and Brossart, (2003), effect sizes of .45 and above may be considered evidence of an effective intervention using the Center M + T technique. Based on this parameter, 6 of the 26 students showed significant differences in oral reading fluency which may be attributed to the Read Naturally intervention when partialling out overall data trend. Note that these students' effect sizes were also considered large using the M + T technique. Also observe that fewer effect sizes were large using this technique. See Figure 20 for a graph of the effect sizes for all 26 students, resulting from the Center M + Trend technique.

In addition to calculating outcomes of single case studies, it is also possible to calculate effect size for replicated AB graphs. This is analogous to a meta-analysis for different studies. Each subject is treated as one separate study and their effects are averaged to compute an overall effect size of treatment success. This procedure is described in Franklin, Allison, and Gorman (1997). The first step to be accomplished

since the data was pre-extracted was to weight the factors. This was done by dividing each subject's effect size (obtained via Center's M + T technique) by the number of data points. Averaging the weighted effect sizes yields $R^2 = 0.04$. Faith, Allison, and

Table 4.

Effect sizes of change between phases for fluency of 26 students using the Center Mean + Trend model with 85% confidence intervals and standard errors.

R^2	Lower Limit	Upper Limit	SE	Graph ID
0.069	0	0.233	0.033	O
0.077	0	0.246	0.032	T
0.083	0	0.278	0.039	B
0.096	0	0.284	0.032	F
0.099	0	0.314	0.040	S
0.102	0	0.314	0.038	E
0.112	0	0.338	0.040	V
0.126	0	0.354	0.035	C
0.154	0	0.415	0.037	Q
0.184	0	0.448	0.028	D
0.194	0	0.423	0.012	Z
0.233	0	0.505	0.014	K
0.247	0	0.490	-0.001	J
0.258	0	0.547	0.011	N
0.267	0	0.556	0.008	U
0.273	0	0.546	0	P
0.385	0	0.652	-0.041	G
0.398	0	0.677	-0.041	Y
0.401	0	0.684	-0.041	W
0.440	0.026	0.664	-0.066	L
0.512	0	0.742	-0.098	M
0.532	0.048	0.742	-0.095	H
0.574	0	0.795	-0.123	R
0.599	0.022	0.796	-0.132	X
0.771	0	0.903	-0.222	I
0.798	0.142	0.912	-0.188	A

Note: For ease of interpretation, the standard error (SE) is an average because the asymmetrical confidence interval is calculated with two SEs. This is a peculiarity of the F distribution. Six students met Parker and Brossart's (2003) criteria of $R^2 = 0.45$ or above for a large effect size; however, two of these students had trendlines in the negative direction. The effect size indicates the amount of variance that can accounted for by the intervention (i.e., Read Naturally). For example, for the student with an effect size of 0.599, it can be stated that 59.9% of the variance in this student's oral reading fluency scores can be explained by the effects of the Read Naturally intervention.

Gorman (1997) also advocate for including confidence intervals around the average weighted mean. This is to account for the amount of variability. For all analyses of confidence intervals in this paper, results obtained with negative values were converted to zero, since effect sizes are not reported in negative numbers. This is computed by the formula: $d \pm (1.96)(\sqrt{\sigma^2}) = 0.04 \pm (1.96)(\sqrt{.001}) = 0$ and 0.102. The authors also recommend reporting the straight arithmetic mean in addition to the weighted mean. This score was 0.307. See Figure 19 for a graph of the effect sizes for fluency using Center M+T with 85% confidence limits.

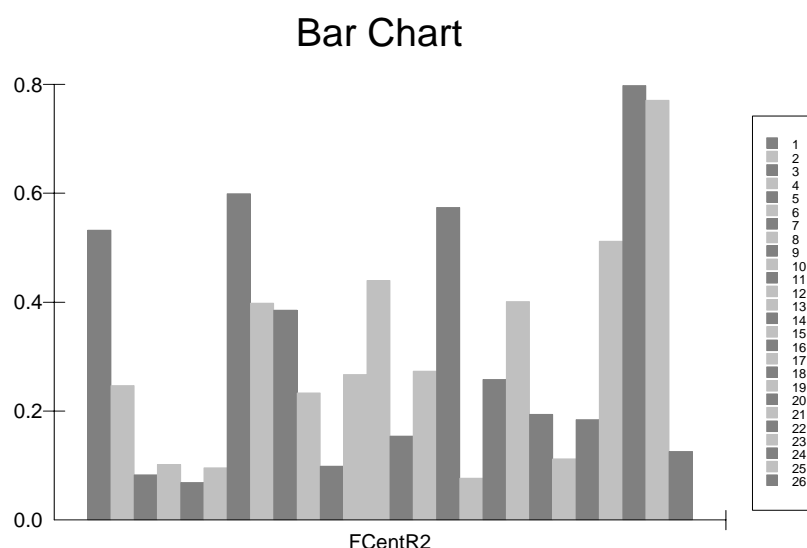


Figure 19. Multiple regression graph of the effect sizes resulting from computation of time series data for the fluency measurement with 85% confidence limits, using Center Mean + Trend. Although effect sizes are only reported in positive numbers, negative numbers were included on this graph to show symmetry. Most effect sizes fell below the 0.45 cutoff that indicates improvement in oral reading fluency.

Including confidence intervals using a variance of 0.05 results in the effect size falling between 0 and 0.738. According to results reported by Parker and Brossart (2003), effect sizes of approximately 0.45 and above are large enough to be indicative of an effective intervention, using the Center M + T technique. Based on this finding, the effect size obtained by combining the data from 26 replicated AB studies is not large, meaning that the intervention does not appear to have been effective for this group of students. Note, however, that there is a low level of certainty about this effect size, based on the confidence interval around the straight arithmetic mean of the fluency effect sizes.

Accuracy skills. In addition to students' oral reading fluency, improvement in accuracy, or the percentage of words read correctly, was assessed. To answer this question, the same techniques that were used to answer the question about fluency were employed. Visual inspection of the raw data graphs of all 26 students revealed less bounce than was present with the fluency graphs. This makes sense when considering that accuracy tends to be less effected by factors such as fatigue because time is not a factor in the accuracy measurement, as it is in the fluency measurement. See Figure 20 for an example of an accuracy graph with low variability (i.e., little "bounce") and one with high variability (i.e., a lot of "bounce").

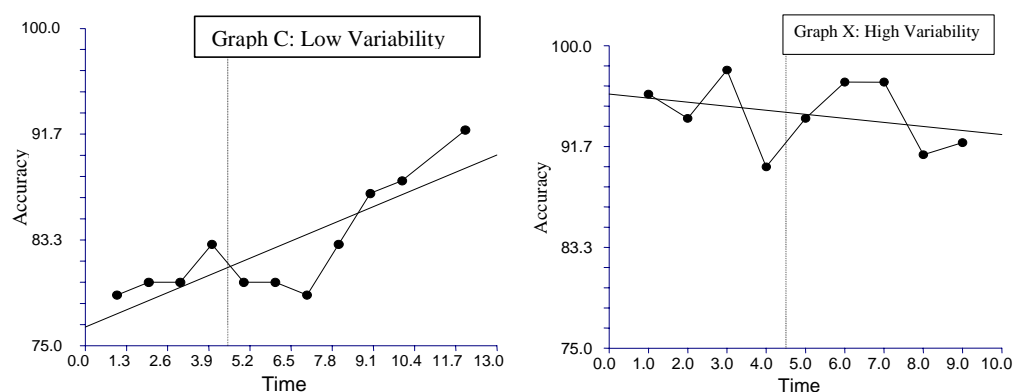


Figure 20. Accuracy graphs depicting low variability (left) and high variability (right).

Graphs are identified by letter for comparison with the complete set of graphs in Appendix C. When taking into account the factors described previously (i.e., bounce, slope, mean, and point of intercept), it appears that 8 of the 26 students showed improvement in their accuracy scores (graphs C, G, H, J, K, N, P, V). Visual analysis findings for accuracy were also corroborated by at least one other professional experienced in visual analysis techniques.

To supplement the visual analysis, the Simple Mean Difference and M + T models were used. Results for the Simple Mean Difference model are reported in Table 5. As with the fluency analyses, autocorrelation was examined and data was cleansed using the ARIMA technique. According to the results reported in the table, 5 students had low to moderate effect sizes (i.e., $R^2 = .30$ and above), indicating that students' mean accuracy scores improved after implementation of the intervention. Of these, 2 students' mean levels decreased after implementation of the intervention, leaving 3 students with improved mean levels post-intervention. After cleansing the data, there were 6 students with low to moderate

effect sizes. Two students had mean levels that decreased post-intervention, leaving 4 students with low to moderate effect sizes.

Table 5.

Effect sizes pre- and post-cleansing of accuracy for 26 students using Simple Mean Difference model with 85% confidence intervals.

Mean R^2	Lower Limit	Upper Limit	SE	R^2 Post-Cleansing	GRAPH ID
0.003	0	0.094	0.031	0.013	X
0.004 ^a	0	0.119	0.039	0	J
0.008	0	0.203	0.065	0.025	T
0.011 ^a	0	0.229	0.072	0.034	M
0.013	0	0.245	0.076	0	O
0.026 ^a	0	0.346	0.102	0.145	Y
0.056 ^a	0	0.412	0.104	0.053	N
0.065	0	0.373	0.084	0.106	H
0.075 ^a	0	0.498	0.121	0.003	I
0.078	0	0.451	0.102	0.175	U
0.079	0	0.38	0.077	0.078	E
0.129 ^a	0	0.544	0.099	0.346	R
0.147 ^a	0	0.591	0.103	0.114	W
0.148	0	0.539	0.084	0.206	V
0.151 ^a	0	0.502	0.069	0.262	B
0.157 ^a	0	0.508	0.067	0.083	P
0.167	0	0.518	0.064	0.197	D
0.178	0	0.498	0.049	0.271	C
0.229	0	0.636	0.062	0.424	F
0.280 ^a	0.009	0.578	0.009	0.295	L
0.298	0	0.646	0.017	0.366	G
0.304	0	0.634	0.009	0.279	K
0.313	0	0.674	0.017	0.514	Q
0.324 ^a	0	0.702	0.019	0.152	A
0.398 ^a	0.029	0.699	-0.024	0.423	S
0.425	0	0.804	-0.016	0.448	Z

Note: For ease of interpretation, the standard error (SE) is an average because the asymmetrical confidence interval is calculated with two SEs. This is a peculiarity of the F distribution. Graph IDs are included to compare with graphs in Appendix C.

^a denotes subject whose mean level decreased from Phase A to Phase B

Taking into account trend, 20 of the 26 students had low to moderate effect sizes (i.e., $R^2 = .30 - .80$; see Table 6), with 3 of these being in the large range ($R^2 = .80$ and above). This indicates that the intervention was effective for improving the accuracy levels of these students. When reflecting back on the results of visual analysis, it was discovered that 8 of the 20 students who had large effect sizes using M + T Difference, actually had results in the negative direction, leaving 12 students with positive effect sizes. See Appendix D for graphs of accuracy scores using the Simple Mean Difference and M + T Difference models for all 26 students.

As with the fluency measure, the stability of the baseline slope was tested for significance. This resulted in only one statistically significant result, precluding use of the Allison M + T technique. Therefore, Center M + T was used for the accuracy analysis. Results are reported in Table 7.

Table 6.

Effect sizes and significance levels of accuracy for 26 students using Mean + Trend

Difference model with 85% confidence intervals.

Mean Plus Trend R^2	Lower Limit	Upper Limit	SE	GRAPH ID
0.089	0	0.305	0.044	H
0.103	0	0.319	0.039	J
0.240	0	0.525	0.016	B
0.244	0	0.512	0.008	E
0.266	0	0.56	0.010	O
0.270 ^{b,c}	0	0.579	0.014	F
0.303 ^{b,c}	0	0.583	-0.008	X
0.317	0	0.61	-0.008	Y
0.374	0	0.654	-0.033	G
0.375	0	0.646	-0.036	K
0.410	0	0.672	-0.051	S
0.412 ^{b,c}	0	0.691	-0.046	U
0.440	0.018	0.673	-0.066	L
0.467	0	0.729	-0.071	Q
0.487 ^b	0	0.726	-0.086	P
0.503	0	0.76	-0.085	R
0.592	0	0.806	-0.131	T
0.611	0	0.817	-0.141	N
0.673 ^b	0	0.854	-0.175	W
0.653	0	0.853	-0.157	Z
0.656 ^{b,c}	0	0.841	-0.164	V
0.667	0.148	0.836	-0.122	D
0.749 ^b	0.222	0.884	-0.136	M
0.837 ^b	0.219	0.933	-0.181	A
0.863 ^b	0	0.614	-0.386	I
0.913	0.734	0.958	-0.047	C

Note: For ease of interpretation, the standard error (SE) is an average because the asymmetrical confidence interval is calculated with two SEs. This is a peculiarity of the F distribution. Graph IDs are included to compare with graphs in Appendix C.

^b denotes subject whose Phase B trend was actually in the negative direction.

^c Subjects with positive mean differences, but negative trends are anomalies. It is often recommended that M + T analysis is not computed in these cases as interpretation is not straightforward.

Table 7.

Effect sizes of accuracy for 26 students using the Center Mean + Trend mode with 85% confidence intervals and standard errors.

R^2	Lower Limit	Upper Limit	SE	GRAPH ID
0.021	0.048	0.742	0.260	H
0.041	0	0.314	0.081	E
0.041	0	0.423	0.118	Z
0.049	0	0.278	0.063	B
0.090	0	0.505	0.113	K
0.103	0	0.49	0.099	J
0.152	0	0.652	0.121	G
0.158	0	0.546	0.080	P
0.174	0	0.314	-0.012	S
0.185	0	0.284	-0.030	F
0.186	0	0.415	0.015	Q
0.221	0	0.795	0.123	R
0.225	0.022	0.796	0.128	X
0.258	0	0.233	-0.098	O
0.309	0	0.677	0.020	Y
0.371	0	0.684	-0.020	W
0.403	0	0.338	-0.163	V
0.409	0.026	0.664	-0.044	N
0.411	0	0.556	-0.092	U
0.428	0	0.547	-0.107	L
0.542	0	0.448	-0.221	D
0.574	0	0.246	-0.313	T
0.716	0	0.354	-0.374	C
0.749	0	0.742	-0.263	M
0.751	0.142	0.912	-0.156	A
0.861	0	0.903	-0.284	I

Note: For ease of interpretation, the standard error (SE) is an average because the asymmetrical confidence interval is calculated with two SEs. This is a peculiarity of the F distribution. Partialling out data trend from both phases left 6 students with improvement in their accuracy scores using Parker and Brossart's (2003) criteria for large effect sizes using the Center Mean Plus Trend technique.

When taking into account mean and partialling out the overall trend, 6 out of 26 students had moderate to large effect sizes, indicating that the intervention was effective for improving their levels of oral reading accuracy, or percentage of words read correctly. Interesting to note is that 3 students (C, D, and T) showed large improvements in their accuracy scores, even though they did not show large improvements in their fluency scores.

The same procedures were followed to compute an overall effect size for accuracy as was used for fluency. Each subject's accuracy R^2 was weighted by dividing the effect size by the total number of data points available. These weighted effect sizes were then averaged for the entire group of 26 students, resulting in $R^2 = 0.041$. See Figure 21 for a graph of the effect sizes with 85% confidence limits, using the Center M + T technique. Most of the effect sizes appear to fall well below the threshold of 0.45, indicating that most students did not show improvement in their accuracy scores.

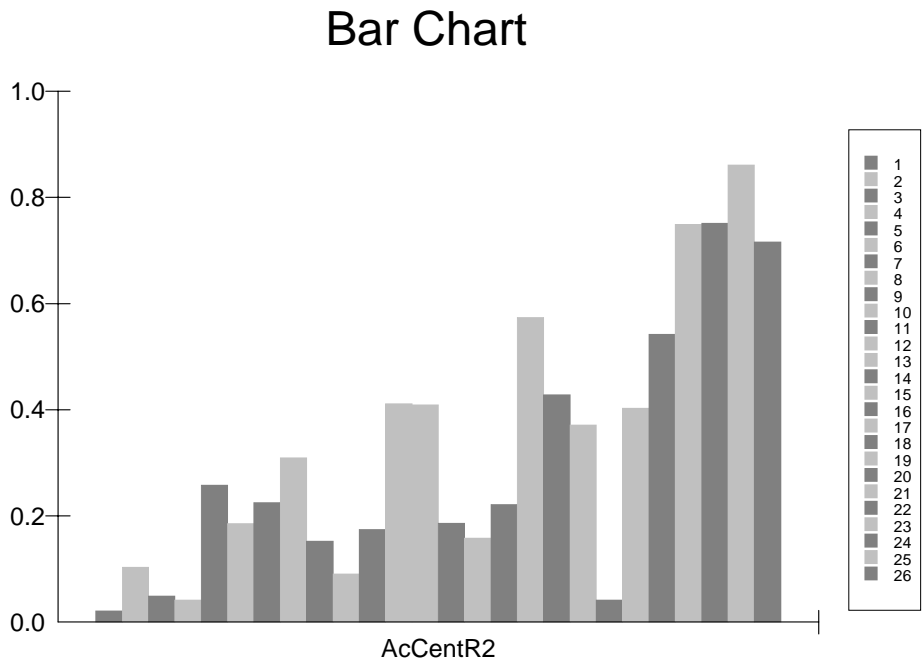


Figure 21. Multiple regression graph of the effect sizes resulting from computation of time series data for the accuracy measurement with 85% confidence limits, using Center Mean + Trend.

Confidence intervals were computed using a variance of 0.001, giving a “true” effect size between 0 and 0.103. This estimated true effect size is much less than the 0.45, considered “large” by Parker and Brossart (2003). Note that this result is similar to the result obtained for fluency. The arithmetic mean for the accuracy effect sizes was $R^2 = 0.324$. The resulting confidence interval using a variance of 0.06 was 0 and 0.804. Again, there was a low level of certainty about this effect size, based on the confidence interval around the straight arithmetic mean of the accuracy effect sizes.

Cross-correlation. Because one of the keys to a successful reading fluency intervention is intensity of treatment, cross-correlations were performed for each subject to determine whether student absences were correlated with their performance on the weekly assessments. Absences were coded as a dichotomous variable. If a student was absent at all during the week of instruction, they received a “1”, if they were not absent at all, a “0” was recorded. The disadvantage of this method is that it weights a single absence the same as being absent all week. Results are displayed in a violin plot in Figure 22. As can be seen in the figure, the average correlation for both measures was zero. For the fluency measurement, only 3 of the 26 students had large negative correlations, indicating that as their number of absences increased, their scores decreased. For the accuracy measurement, 1 student had a large negative correlation.

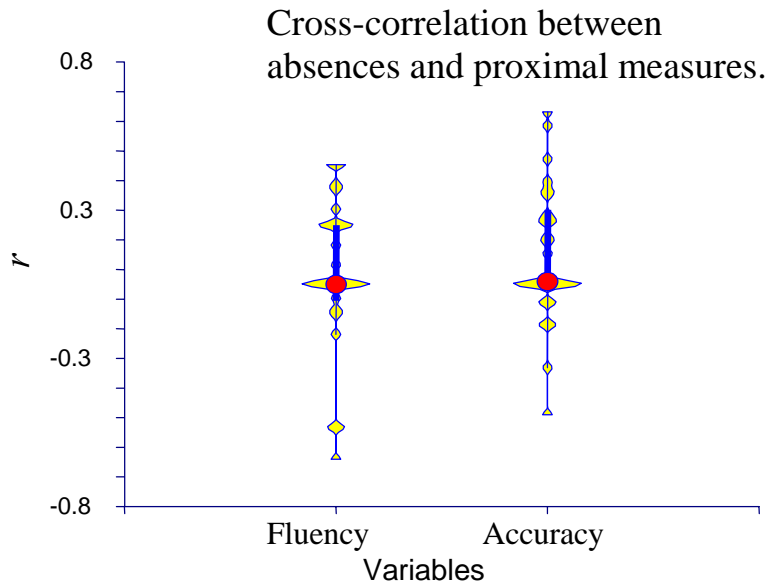


Figure 22. Violin plot of cross-correlations (lag-1) of absences with fluency and accuracy scores.

Distal Measures

To supplement the time series data, distal measures of the students' fluency and accuracy skills were included. Each measure was taken during the baseline phase and again around the students' 30th day in the study, using alternate forms. Results were considered statistically significant if $p < .05$. Means and standard deviations of scores pre and post intervention are reported in Table 8.

Table 8.

Means and standard deviations for subscales measuring fluency and accuracy.

Scale	Pre-Test		Post-Test	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
TOWRE Sight Word	70.33	10.32	70.67	9.92
TOWRE Phonetic Decoding	65.73	10.85	68.67	10.56
TOWRE Word Reading Efficiency	61.65	11.43	63.50	11.27
WRMT-R Word Atta	77.31	10.31	79.17	8.61
GORT-III Rate	1.52	0.92	1.88	1.26
GORT-III Accuracy	2.36	1.70	2.50	2.32
GORT-III Oral Reading Quotient	65.8	16.25	68.88	17.71

Note: For the TOWRE, WRMT-R, and GORT-III Oral Reading Quotient scales, $X = 100$, $SE = 15$. For the GORT-III scales, $X = 10$, $SE = 3$.

Note that there were improvements in the students' mean scores on all measures between pre- and post-test. It may seem surprising that students' mean score was higher on the WRMT-R Word Attack subtest than on the TOWRE Phonetic Decoding subtest since they both consist of similar tasks. However, the TOWRE is timed and the WRMT-R subtest is not, which may account for the lower scores on the TOWRE subtest.

The Test of Word Reading Efficiency (TOWRE) included two scales, as well as an overall scale score. The scales all had a mean of 100 and standard deviation of 15.

Repeated measures within-subject Analysis of Variance was conducted for each of these

scales. A total of 23 students were included for the analyses as three students returned to their home schools prior to post-testing.

The TOWRE Sight Word scale required students to read as many words as they could from a list in 45 seconds. Because of the time component combined with the accuracy component, this task may be considered a test of oral reading fluency. Students did not show significant improvement on this task post-intervention, $F = 0.18$, $p = 0.675$, $R^2 = 0.04$. The TOWRE Phonetic Decoding scale required students to read nonsense words phonetically. This task also had a time component and may be considered a measure of students' oral reading fluency of phonetically spelled nonsense words. Students showed statistically significant improvement on this task post-intervention, $F = 13.26$, $p = 0.001$, $R^2 = 0.05$. In addition to the two subscales, the TOWRE provides an overall Word Reading Efficiency score, which takes into account students' performance on the two subscales. For this scale, students showed significant improvement, $F = 5.66$, $p = 0.026$, $R^2 = 0.04$.

The Word Attack subtest of the Woodcock Reading Mastery Test-Revised (WRMT-R) was included as an additional measure of students' reading accuracy. Students were required to read a list of nonsense words, sounding them out phonetically. There was no time limit on this test; therefore, it is a test of accuracy rather than fluency. Students' scores showed statistically significant improvement on this subtest subsequent to the intervention, $F = 5.13$, $p = 0.033$, $R^2 = 0.09$.

The Gray Oral Reading Tests, Third Edition (GORT-III) was included to supplement the findings of the TOWRE and WRMT-R. The GORT-III included

measures of rate (fluency) and accuracy, as well as an overall score that took into account these two scales plus a comprehension component. ANOVA was performed using the scores from 22 students for all the GORT-III scales. Three students were unavailable at the post-test as previously mentioned and one student's lack of cooperation on the pretest rendered his results invalid. The result of analysis of students' scores on the Rate scale was borderline, $F = 4.04$, $p = 0.057$, $R^2 = 0.21$. In contrast, the students did not show significant improvement on the Accuracy scale, $F = 0.46$, $p = 0.505$, $R^2 = 0.17$. The GORT-III Oral Reading Quotient took into account the students' scores on these scales, plus a comprehension measure (to be discussed in the next section). Students' did not show significant improvement, $F = 1.33$, $p = 0.261$, $R^2 = 0.24$.

Does the reading intervention have an impact on reading comprehension of the students? To answer this question, students were given the Passage Comprehension subtest of the Woodcock Reading Mastery Tests-Revised (WRMT-R). Students read passages silently and were required to orally fill in the blank with an appropriate word. Students were given as much time as they needed to complete each item. Scores from 23 students were included. On the pre-test assessment, $X = 65.38$, $SD = 14.97$. On the post-test assessment, $X = 69.75$, $SD = 11.09$. Subsequent to intervention, students' improvement was statistically significant, $F = 5.64$, $p = 0.026$, $R^2 = 0.11$. This is a surprising finding, considering that comprehension is very difficult to improve in a short period of time.

In addition, there was a comprehension component to the Gray Oral Reading Tests, Third Edition (GORT-III). Students read passages then answered five questions about each

passage. Scores from 22 students were included, due to unavailability of three students and invalid results from a fourth student. On the pre-test assessment, $X = 2.80$, $SD = 2.25$. On the post-test assessment, $X = 3.13$, $SD = 1.90$. Students scores post-intervention were not statistically significant, $F = 2.10$, $p = 0.162$, $R^2 = 0.45$.

How does this intervention affect the students' self-efficacy with regards to reading? In addition to students' reading skills, the researcher was interested in exploring whether the reading intervention may have had an impact on students' self-efficacy toward reading. The Self-Rating of Reading scale, devised by the researcher, contained 6 questions pertaining to how students felt about their reading rate, accuracy, and comprehension. The students rated their opinions on a 4-point Likert-type scale. Scores were added for all 6 questions to obtain an overall Self-Rating of Reading Score, with a minimum score of 6 and maximum score of 24 possible. The mean for this scale was 15, with a standard deviation of 1. The scores of 23 students were included as three students had returned to their home campus prior to post-testing. For the pre-intervention assessment, $X = 9.73$, $SD = 2.76$. Post-intervention, $X = 10.92$, $SD = 3.93$. Repeated measures within subjects ANOVA resulted in $F = 7.11$, $p = 0.014$, $R^2 = 0.14$. This result is statistically significant at $p < 0.05$.

How does the intervention affect the way students compare themselves to others with regards to reading?

In addition to questions regarding students' opinions of their reading skills, the Self-Rating of Reading scale included a 10-point scale asking students to rate how well they read compared to same-aged peers in their home schools, given anchors. Twenty-

three students were included in this assessment. On the pre-test, $X = 1.77$, $SD = 0.99$.

Post-test, $X = 2.33$, $SD = 1.09$. ANOVA resulted in $F = 10.27$, $p = 0.004$, $R^2 = 0.28$.

This result is statistically significant at $p < 0.05$.

Supplementary Informal Analysis

In addition to providing visual and statistical data, one of the benefits of single case design is that one is able to examine results at an individual level. After examining the quantitative data, the results were studied qualitatively to determine if there were any visible patterns to explain why some students improved their fluency and/or accuracy scores and some did not. Because Type II error (failing to identify improvement) is of more concern than Type I error (identifying improvement that is not present) is more of a concern for this study, as discussed in Chapter III, the less conservative method of Mean + Trend was utilized for comparison.

Data related to gender, ethnicity, absences, instructors, age, and grade level, were examined informally to identify patterns. Upon studying the data, it was determined that information related to ethnicity, grade, and age were not informative as it pertained to improvement. Therefore, information related to gender, participation in special education, teacher, initial reading level, and absences were included in Table 9, along with effect sizes for fluency and accuracy using the Mean + Trend difference model.

Table 9.

Qualitative information and effect sizes using the Mean + Trend technique for each student.

Student	Gender	Special Education	Initial Reading Level	Teacher	Weeks With At Least One Absence	Total Weeks In Study	Fluency M + T R^2	Accuracy M + T R^2
F	Male	Yes	6.5	A	2	7	0.097 ^b	0.270 ^{b,c}
V	Male	Yes	2.0	B	2	8	0.113	0.656 ^{b,c}
B	Male	Yes	0.8	A	0	10	0.141	0.240
P	Male	Yes	5.0	B	2	10	0.276	0.487 ^b
T	Female	Yes	4.0	A	3	8	0.295 ^b	0.592
E	Male	Yes	2.5	B	2	12	0.313	0.244
U	Female	Yes	4.0	B	0	8	0.317 ^b	0.412 ^{b,c}
O	Male	No	4.0	A	1	9	0.350	0.266
K	Male	No	3.0	C	5	10	0.352	0.375
Z	Male	No	3.0	A	3	5	0.362	0.653
S	Male	No	4.5	A	8	10	0.411	0.410
Y	Female	Yes	2.5	A	0	7	0.435 ^b	0.317
L	Male	No	1.5	C	3	13	0.437	0.440
C	Male	Yes	2.5	B	6	12	0.460	0.913
W	Female	Yes	3.0	A	0	6	0.487 ^b	0.673 ^b
M	Female	Yes	1.5	C	0	9	0.508	0.749 ^b
J	Male	No	2.0	C	5	14	0.524	0.103
D	Male	No	3.0	A	4	10	0.527	0.667
H	Male	Yes	0.8	B	6	11	0.546 ^b	0.089
Q	Male	Yes	6.0	A	7	8	0.561	0.467
X	Male	Yes	5.0	A	5	9	0.631 ^b	0.303 ^{b,c}
N	Male	Yes	2.5	C	2	8	0.692	0.611
G	Male	Yes	2.0	B	4	9	0.752	0.374
R	Female	No	1.0	A	0	7	0.814	0.503
I	Female	No	4.0	A	0	6	0.817 ^b	0.863 ^b
A	Male	No	2.0	A	1	7	0.859	0.837 ^b

^b denotes subject whose Phase B trend was actually in the negative direction.

^c Subjects with positive mean differences, but negative trends are anomalies. It is often recommended that M + T analysis is not computed in these cases as interpretation is not straightforward.

Improvement was considered to be $R^2 = .30$ or greater, and trendlines were required to be in the positive direction. As can be seen from the table, 13 out of 19 (68%) males showed improvement on fluency and 9 out of 19 (47%) males showed improvement on accuracy. Two out of 7 (29%) females showed improvement on fluency and 3 out of 7 (43%) on accuracy. Six out of 15 (40%) of students showing improvement in fluency were in special education classes, while 10 out of 12 (83%) students showing improvement in accuracy were in special education classes. Based on informal checks of fidelity of program implementation, Teacher C tended to show the most consistency and intensity of implementation, while Teacher B required the most reteaching and monitoring. Teacher A taught most of the students, and this teacher had an even split of 7 students showing improvement in fluency and accuracy, and 7 showing no improvement in both. Teacher B taught 7 of the students, with 3 showing improvement in fluency and 2 showing improvement in accuracy. Teacher C taught 5 students, and all 5 showed improvement in fluency, while 3 showed improvement in accuracy. There were positive results for students who were reading at least six levels below grade level, as well as for those who were “only” two levels behind.

As discussed previously in the section about cross-correlation, absences were determined by indicating whether a student was absent at all during the week before CBM. This was done because there was only one CBM measurement per week and to do cross-correlation, there had to be an equal number of measures of absences as CBMs. Therefore, a student that was absent 5 times in one week received the same score as a student who was absent one time. The distribution of absences for students who

improved and those who did not was similar. Only 6 students improving on fluency were absent less than 25% of the time, while 9 were absent 25% of the time or more. For students not improving on fluency, 6 were absent less than 25% of the time, while 6 were absent more. For accuracy, 4 students showing improvement were absent less than 25% of the time, while 8 were absent more than 25% of the time. Nine students who did not show improvement on accuracy were absent less than 25% of the time, while 6 students were absent more than 25% of the time.

CHAPTER V

DISCUSSION AND SUMMARY

Prior to discussing the results, several limitations of the study are addressed. One limitation of this study was that it did not allow us to judge whether the students' acquired skills transferred to learning in their home schools upon returning. In addition, although it is relatively easy to detect growth with regard to oral reading fluency and accuracy (Hasbrouck, Woldbeck, Ihnot, & Parker, 1999), comprehension improvement is more difficult to detect in the short-term. This is because comprehension is a highly complex skill (LaBerge & Samuels, 1974) and is not easily assessed using Curriculum-based Measurement (CBM). Therefore, it is possible that a student's fluency may have improved, but comprehension did not. Another limitation was that, although the teachers have been formally trained in the Read Naturally (RN) strategy, three different teachers delivered instruction to their assigned students. As is the case in many situations, some of the teachers were more adept at implementing and monitoring the intervention than others. Although checks were made by the researcher to ensure the fidelity of instruction and extra support was provided when needed, it is possible that some students received more intensive instruction than others. Finally, lack of motivation may have influenced the scores of some students. Research has shown that motivation is an important component of learning to read fluently (Ihnot, 2001). While motivation was not formally measured, some students were clearly excited about participating in the

program as indicated by their verbalizations and involvement, while others required constant prompting by the instructors to stay on-task and complete the tasks.

The results of this study were mixed as to whether an intensive, short-term reading intervention targeting oral reading fluency can be effective for improving adolescent students' oral reading fluency, accuracy, comprehension, and self-efficacy toward reading. Each research question will be discussed in turn.

Does a Short-Term, Intensive Reading Intervention Targeting Oral Reading Fluency Improve the Reading Accuracy and Reading Fluency Skills of Students Temporarily Assigned to an Alternative School Due to Behavior Problems?

The answer to this question was yes and no. Although significant improvement was not seen for an overwhelming majority of students on the time series measure (i.e., curriculum-based measurement), some students did improve their fluency and accuracy. Using the more stringent results from the Center Mean Plus Trend model, only 6 of the 26 students improved their oral reading fluency based on weekly one-minute curriculum-based measurements (CBM). Results for the accuracy measure (i.e., percentage of words spelled correctly) were similar, with 6 of the 26 students showing significant positive results. However, as discussed in previous chapters, the consequences for falsely identifying positive results is not as detrimental as failing to see results in a program due to false negatives. Therefore, taking into consideration that Type II error is preferable to Type I error, the less conservative method of Mean + Trend was considered sufficient for determining improvement. When using this method, 15 out of 26 students showed significant improvement on fluency and 12 out of 26 students showed

significant improvement on accuracy. This is evidence that for a large amount of students in the disciplinary alternative placement, oral reading fluency and accuracy can improve, despite a relatively short-term placement.

Looking at the distal measures, the students did not show statistically significant improvement on fluency and accuracy as measured by the GORT-III. However, although students taken as a group did not significantly improve their sight-word reading as measured by the TOWRE Sight Word subtest, they did evidence statistically significant improvement in phonetic decoding and their overall fluency score on this measure. In addition, accuracy as measured by the WRMT-R Word Attack subtest improved as well. These results lend validity to the positive results obtained on the time series measure.

There are several reasons why students overall may have shown positive results on these measures and not on the CBM. One reason is that students' baselines were very short, which contributed to the amount of variability in their scores. Without a stable baseline, it is very difficult to show significant improvement. Compounding this issue, most students also had very short intervention phases. This occurred for various reasons. As discussed next, absences were a problem for some students. Even if a student was present for instruction during most of the week, it was often the case that they were absent on the day of the weekly CBM. Some students were discharged early for good behavior.

Although some students improved their oral reading fluency and accuracy, it appeared that some students' performance actually deteriorated; however, there may be

contributing factors. As indicated in the Supplementary Analysis section of Chapter IV, there were no clear-cut characteristics of students who improved. In general, students who were male, did not participate in special education classes, and were taught by Teacher C appeared to be the most successful. For some students, repeated absences, including suspensions were problematic. Due to the way data was collected for the time series measure (i.e., once per week), cross-correlations were conducted considering whether an absence occurred during the week at all. Had the total number of absences been considered, it might have been shown that this was a major interference with students' performance. For example, some students were suspended or expelled for more than a week, yet this was recorded as the same as a student who was absent only one time in a week. However, there was an approximately equal distribution of absences for students who improved and those who did not.

It was disappointing that the students' scores did not significantly improve on the fluency and accuracy subtests of the GORT-III. This measure most closely resembled the CBM measure. Because students showed significant improvement on the TOWRE and WRMT-R subtests, it appears that, although they made improvements on word reading in isolation, passage reading remained difficult for some of the students.

Does the Reading Intervention Have an Impact on Reading Comprehension of the Students?

Although comprehension is the ultimate goal of reading, due to the complexity of this skill it was not expected that there would be significant improvement in the short-term (LaBerge & Samuels, 1974). Again, results were mixed. Results on the Passage

Comprehension subtest of the WRMT-R indicated statistically significant improvement in students' comprehension. However the results of the GORT-III were not statistically significant. There may have been a discrepancy for several reasons. The task requirements were slightly different on these measures. The GORT-III required students to read a passage and then answer 5 questions related to the passage. This is typically considered an informal measure of comprehension, due to the limited number of questions one must answer. In addition, as discussed in Chapter II, if students are unable to read a passage fluently (as indicated by the rate subscale), then it is likely that they will not understand what they have read (LaBerge & Samuels, 1974). This appears to be what occurred on this measurement for some students. Although there was not significant improvement on this measure, because students' fluency appeared to improve, it is possible that measurable improvements in comprehension would be seen with continuation of the program.

The WRMT-R required the students to read progressively more difficult passages, ranging from simple sentences to complex paragraphs. A word was missing from each passage, and students were required to fill in the blank orally with the most appropriate word, based on their understanding of the passage. In general, the passages were much shorter than those required of the GORT-III. In addition, students were not timed on this task, which allowed them to take as much time as they wished to read and reread each word. It is possible that simply removing the stopwatch helped some students feel more at ease. Perhaps some students felt less overwhelmed by the relatively shorter passages.

How Does This Intervention Affect the Students' Self-Efficacy With Regards to Reading?

Perhaps even more important than actual improvement in oral reading fluency, accuracy, and comprehension was whether the students' perceptions about their ability to read improved. Although the other skills are certainly important, if students' self-efficacy toward reading could improve despite their actual scores, it is possible that their motivation and attitudes can also be effected, which could lead to more effort and eventual improvement in these skills (Bandura, 1997).

As reported in the previous chapter, self-efficacy toward reading appeared to improve significantly, as measured by the Self-Rating of Reading measure. This result is an important reflection of the improvement students showed on the reading measures. There may be several reasons for the improvement in students' improved confidence. Due to the nature of RN and CBM, students' success was based on where they began. Achievement was based on each student's individual goals, not comparison with others. Because students compared their progress to themselves, each student was able to see some success. Each student graphed the results of their improvement on each story read during instruction, and they were required to "pass" a story based on individual goals before moving to the next. Therefore, each student was able to feel successful, even if his or her overall scores did not improve significantly. In addition, although students worked fairly independently, they received individual attention from teachers. It is possible that individual feedback helped improve the students' self-efficacy toward reading. Because of the importance of self-efficacy (Bandura, 1977), this finding may have significant implications on these students' future academic growth. In general,

students were beginning to feel more positive about their abilities to read. As discussed previously, a link has been shown between disruptive behavior and reading difficulties (e.g., Dishion, Loeber, Southamer-Loeber, and Patterson, 1984). Perhaps further study would have shown increased positive behaviors for students demonstrating improvement in their oral reading fluency, accuracy, and self-efficacy.

How Does the Intervention Affect the Way Students Compare Themselves to Others With Regards to Reading?

Bear, Minke, Griffin, and Deemer (1998) discussed research indicating that students with learning disabilities were able to realistically assess their abilities, rating themselves more negatively than their peers. Although they also note the importance of social comparison, stating that multiple failures and frustrations can lead to more dissatisfaction and less motivation toward reading. Therefore, it was important that students begin to rate themselves more similar to their peers as they began to see improvement in their oral reading fluency and accuracy.

Students showed statistically significant improvement in the way they viewed their reading skills compared to same-age peers. On this scale, students were asked to compare their reading skills (e.g., fluency, accuracy, and comprehension) to students in their home schools. Although students tended to rate themselves lower than their peers, by the end of the intervention, they rated themselves closer to their peers. This is a promising finding that supports the self-efficacy measure. Not only did students feel better about their own abilities, they did not feel that they were as far behind their peers

with regard to reading. This is an important finding that may have implications for the students' continued success.

Summary and Implications for Future Research

The results of this study were promising. Although some measures appeared to show improvement in reading fluency, accuracy and comprehension, others did not. Using the less conservative time-series method of analysis, more than half of the students showed improvement on oral reading fluency and slightly less than half showed improvement in accuracy. This study showed that, under some circumstances, students can improve their reading skills significantly, even during a very short placement at a disciplinary alternative school setting. Students also showed improvement in their self-efficacy toward reading and social comparison with regard to reading. This is an important finding, since one of the contributions of Read Naturally as a program is that it has the component of the students graphing their progress, with the purpose of increasing motivation (Ihnot, 2001). Although as a group the students did not significantly improve their fluency and accuracy, this study was for a very short duration. Therefore, this may be evidence that, if continued, the program may be a precursor to improvement in reading, since it appears to have positive effects on students' self-efficacy toward reading and comparison with others.

As noted in the supplementary informal analysis in Chapter IV, there was not a clear-cut explanation for why some students evidenced significant improvement in their oral reading fluency and/or accuracy and others did not. The students who improved were of various ethnicities, ages, grade levels, reading levels, and classrooms. In

addition, they were males and females and had varying rates of absenteeism. The same was true for students whose graphs tended to decrease. One possible explanation for this, as mentioned earlier, is that there were other factors that influenced the students' performance such as motivation or being "sold" on the program. Perhaps level of involvement with their lessons also played a role. It was interesting to note that the teacher (i.e., Teacher C) who seemed to be the most consistent and cognizant of the program, had 5 out of 5 students show improvement in fluency and 3 out of 5 show improvement in accuracy. This suggests that the more intense and consistent the implementation, the more successful the program can be, even on a short-term basis.

The results of this study show promise for future research with reading interventions for students in short-term alternative education placements. To date, there exists mainly anecdotal evidence for the success of disciplinary alternative education placements (Cortez and Robledo Montecel, 1999). This study offers evidence that short-term disciplinary alternative education placements can provide intensive, successful programs to improve the oral reading fluency, accuracy, comprehension, self-efficacy toward reading, and social comparison of students struggling with reading upon placement. Perhaps remediation of reading skills prior to placement would reduce frustration levels and improve student behavior, reducing the need for referral to disciplinary alternative education campuses.

There are several things that could be improved and expanded upon for future research. Although the teachers were trained in the RN method prior to implementation, some teachers' understanding of how to place students at the appropriate reading level

for instruction required remediation during the study. Therefore, more intense training and assessment of the teachers' understanding and motivation may have resulted in more positive results. In addition, teachers often waited to place students in the program until they were at the school for several days. This hesitation resulted in less instruction and fewer time-series datapoints. Therefore, a more efficient system of placing students would lend to more positive results.

In addition to improving the training of teachers and methodology, there are several questions that were not answered by this study. It would be important to measure students' motivation to participate in the program. Negative results would be interpreted very differently for a student who appeared to be highly motivated and exerted a lot of effort compared to a student who was not motivated to participate and gave only superficial effort.

As briefly discussed in Chapter II, research has pointed to a link between behavior and reading (Frick, et al., 1991). Because students were placed in the alternative education setting due to their negative behaviors at their home schools, it would be important to expand this research to study the effect of improving reading skills on behavior. In addition, follow-up to study generalization into the home campus and/or provide continued instruction after exiting the temporary placement would be important extensions of this study.

Despite the challenges faced by staff and administrators at an alternative education placement, this study provides preliminary evidence that these placements do not need to be the "dumping grounds" described by Cortez and Robledo Montecel

(1995). Rather, these schools are in a unique position to provide intensive reading fluency interventions such as RN in order to intervene and help students begin to feel success.

REFERENCES

- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavior, Research, and Therapy*, 31, 621-631.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York: W.H. Freeman and Company.
- Bandura, A., Barbaranelli, C., Vittorio Caprara, G., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72, 187-206.
- Bear, G. G., Minke, K. M., Griffin, S. M., & Deemer, S. A. (1998). Achievement-related perceptions of children with learning disabilities and normal achievement: Group and developmental differences. *Journal of Learning Disabilities*, 31, 91-104.
- Blatchford, P. (1997). Students' self assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison group. *Educational Psychology*, 17, 345-359.
- Blau, Z. S. (1981). *Black children/white children: Competence, socialization, and social structure*. New York: Free Press.
- Busk, P. L. & Marascuilo, L. A. (1992). Statistical analysis in single-case research: Issues, procedures, and recommendations, with application to multiple behaviors. In T. R. Kratochwill and J. R. Levin (Eds.), *Single-case research design and*

analysis. New directions for psychology and education (pp. 159-185). Mahwah, NJ: Lawrence Erlbaum Associates.

Catterall, J. S. (1985). *On the social costs of dropping out of school*. Palo Alto, CA: Stanford University, Stanford Educational Policy Institute, School of Education.

Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill Book Company.

Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*. 12 (4), 425-434.

Committee for Economic Development. (1987). *Children in need: Investment strategies for the educationally disadvantaged*. New York: Committee for Economic Development.

Cortez, A., & Robledo Montecel, M. (1999). *Disciplinary alternative education programs in Texas—What is known; what is needed. A policy brief*. (ERIC Document Reproduction Service No. ED434963).

Curtis, M. E., & Longo, M. A. (1999). *When adolescents can't read: Methods and materials that work*. Cambridge: Brookline Books.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, 30, 507-524.

Dishion, T. J., Loeber, R., Stouthamer-Loeber, M., & Patterson, G. R. (1984). Skill deficits and male adolescent delinquents. *Journal of Abnormal Child Psychology*, 12, 37-54.

Duckenfield, M. (1990). *Effective strategies for dropout prevention: Twelve successful strategies to consider in a comprehensive dropout prevention program.*

Clemson, SC: National Dropout Prevention Center. (ERIC Document Reproduction Service No. ED322461).

Faith, M. S., Allison, D. B., & Gorman, B. S. (1996). Meta-analysis of single-case research. In R. D. Franklin and D. B. Allison (Eds.), *Design and analysis of single-case research*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Farrington, D. P. (1979). Environmental stress, delinquent behavior, and conviction. In I. G. Sarason & C. D. Spielberger (Eds.), *Stress and anxiety*, 93-107.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.

Franklin, R. D., Allison, D. B., & Gorman, B. S. (1996). *Design and analysis of single-case research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Freeman, G., Gum, M., & Blackburn, J. M. (1999). *Proactive approaches to improving outcomes for at-risk students*. Oxford: University of Mississippi. (ERIC Document Reproduction Service No. ED430948).

Frick, P. J., Kamphaus, R. W., Lahey, B. B., Loeber, R., Christ, M. A. G., Hart, E. L., & Tannenbaum, L. E. (1991). Academic underachievement and the disruptive behavior disorders. *Journal of Consulting and Clinical Psychology*, 59, 289-294.

Fuchs, L. S., Fuchs, D. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*, 22, 27-48.

Gorman, B. S. & Allison, D. B. (1997). Statistical alternatives for single-case design. In R. D. Franklin, D. B. Allison, and B. S. Gorman (Eds.), *Design and analysis of single-case research*. (pp. 159-214). Mahwah, NJ: Lawrence Erlbaum Associates.

Hale, L. F. (2001). School dropout prevention. Information and strategies for parents. *Communiqué*, 29 (6), Insert.

Halsell Miranda, A., & Santos de Barona, M. (1990). A model for interventions with low achieving minority students. In A. Barona and E. E. Garcia (Eds.), *Children at risk: Poverty, minority status, and other issues in educational equity* (pp. 119-134). Washington, DC: The National Association of School Psychologists.

Hasbrouck, J. E., Ihnot, C., & Rogers, G. H. (1999). "Read naturally": A strategy to increase oral reading fluency. *Reading Research and Instruction*, 39, 27-36.

Hasbrouck, J. E., & Tindal, G. (1991). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children*, 24 (3), 41-44.

Hasbrouck, J. E., Woldbeck, T., Ihnot, C., & Parker, R. I. (1999). One teacher's use of curriculum-based measurement: A changed opinion. *Learning Disabilities Research and Practice*, 14 (2), 118-126.

Henk, W. A., & Melnick, S. A. (1998). Upper elementary-aged children's reported perceptions about good readers: A self-efficacy influenced update in transitional literacy contexts. *Reading Research and Instruction*, 38, 57-80.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin*, 111, 127-155.

Hinshaw, S. P., & Anderson, C. A. (1996). Conduct and oppositional defiant disorders. In E. J. Marsh & R. A. Barkley (Eds.), *Child psychopathology* (pp. 113-149). New York: The Guilford Press.

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician* 52, 181-184.

Horn, M. (1987, May 18). The burgeoning educational underclass. *U.S. News and World Report*, 102, 66-67.

Ihnot, C. (2001). *Read naturally*. Retrieved February 12, 2002, from <http://www.readnaturally.com>. St. Paul: Read Naturally.

Information Services. (2000). *Government information sharing project*. Retrieved October 1, 2001, from <http://govinfo.library.orst.edu/cgi-bin/sddb-state?stateis=Te>.

Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology*, 36 (4), 399-415.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.

Lange, C. M. & Sletten, S. J. (2002). *Alternative education: A brief history and research synthesis*. (ERIC Document Reproduction Service No. ED462809).

Madelaine, A., & Wheldall, K. (1999). Curriculum-based measurement of reading: A critical review. *International Journal of Disability, Development & Education*, 46, 71-85.

Marascuilo, L. A. & Busk, P. L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment, 10*, 1-28.

Marston, D & Magnusson, D. (1988). Curriculum-based measurement: District level implementation. In J. Graden, J. Zins, & M. Curtis (Eds.), *Alternative education delivery systems: Enhancing instructional options for all students* (pp.137-172).

Washington, DC: National Association of School Psychologists.

McCreight, C. (1999). *Best practices for disciplinary alternative education programs in Texas*. (ERIC Document Reproduction Service No. 437 715).

National Reading Panel (2000). *Teaching children to read—Summary report*. Retrieved June 15, 2001, from <http://nationalreadingpanel.org/Documents/default.htm>.

National Research Council. (1998). *Preventing reading difficulties in young children*. C. E. Snow, S. Burns, & P. Griffin (Eds.). Washington, DC: National Academy Press.

Parker, R. I. & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy, 34* (2), 189-211.

Parsonson, B. S. & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis. New directions for psychology and education*. (pp.15-40). Mahwah, NJ: Lawrence Erlbaum Associates.

Raywid, M. A. (1994). *Focus on schools: A genre to consider*. New York: (ERIC Document Reproduction Service No. ED377293).

- Reitzammer, A. F. (1991). Dropout prevention: The early years. *Reading Improvement, 28*, 255-256.
- Reutzel, D. R., & Hollingsworth, P. M. (1993). Effects of fluency training on second graders' reading comprehension. *Journal of Educational Research, 86*, 325-331.
- Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health, and behavior*. London: Longman.
- Samuels, S. J. (1986). Why children fail to learn and what to do about it. *Exceptional Children, 53*, 1-16.
- Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly: Overcoming Learning Difficulties, 13*, 107-121.
- Sanders, M. (2001). *Understanding dyslexia and the reading process. A guide for educators and parents*. Boston: Allyn and Bacon.
- Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.
- Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of Educational Psychology, 73*, 93-105.
- Schunk, D.H. (1984). Enhancing self-efficacy and achievement through rewards and goals: Motivational and informational effects. *Journal of Educational Research, 78* (1), 29-34.
- Schunk, D. H., & Rice, J. M. (1989). Learning goals and children's reading comprehension. *Journal of Reading Behavior, 21* (3), 279-293.

- Seligman, M. E., & Maier, S. F. (1967). Failure to escape traumatic shock. *Journal of Experimental Psychology*, 71, 1-9.
- Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology*, 87, 386-398.
- Shinn, M. R. (1989). *Curriculum-based measurement. Assessing special children*. New York: The Guilford Press.
- Southwest Educational Development Laboratory (SEDL). (1995). Alternative learning environments. *Insights on educational policy and practice*. Austin, TX: Author.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 11, 360-406.
- Stanton, W. R., Feehan, M., McGee, R., & Silva, P. A. (1990). The relative value of reading ability and IQ as predictors of teacher-reported behavior problems. *Journal of Learning Disabilities*, 23, 514-517.
- Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed, Inc.
- Wiederholt, J. L., & Bryant, B. R. (1992). *Gray Oral Reading Tests*, Third Edition. Austin, TX: Pro-Ed, Inc.
- Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin*, 106, 231-248.

Woodcock, R.W. (1987). *Woodcock Reading Mastery Tests, Revised*. Circle Pines, MN: American Guidance Service.

Young, T. W. (1990). *Public alternative education: Options and choices for today's schools*. New York: Teachers College Press.

APPENDIX A

SELF RATING OF READING

Examiner CopyExaminer says:

“This is a social studies book similar to what you have in your classroom.”

Flip through the book, showing the student several pages. However, do not spend more than 2 seconds on one page. This prevents the student from actually reading the text to make a judgment.

“Most students your age can read a page from this book in 5 minutes. How sure are you that you can read a page from this book in 5 minutes?”

1. Not Sure Maybe Pretty Sure Real Sure

**Examiner says:**

“Most students your age can read this page with very few errors. How sure are you that you could read this page with only a few errors?”

2. **Not Sure Maybe Pretty Sure Real Sure**



Examiner says:

“If I ask you to read a page from this book then answer 5 questions about how well you understand the story, how sure are you that you would answer all 5 questions right?”

3. **Not Sure** **Maybe** **Pretty Sure** **Real Sure**



Examiner says:

“When you are back in your regular school, how sure are you that you could read your classroom textbook out loud to your class without making a lot of mistakes?”

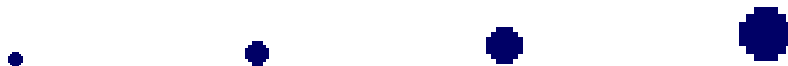
4. **Not Sure** **Maybe** **Pretty Sure** **Real Sure**



Examiner says:

“When you are back in your regular school, how sure are you that you could finish reading a chapter out of your classroom textbook in the same amount of time as most other students?”


5. **Not Sure** **Maybe** **Pretty Sure** **Real Sure**



Examiner says:

“When you are back in your regular school, how sure are you that you can read and understand your classroom textbook?”

6. **Not Sure** **Maybe** **Pretty Sure** **Real Sure**



Social Comparison

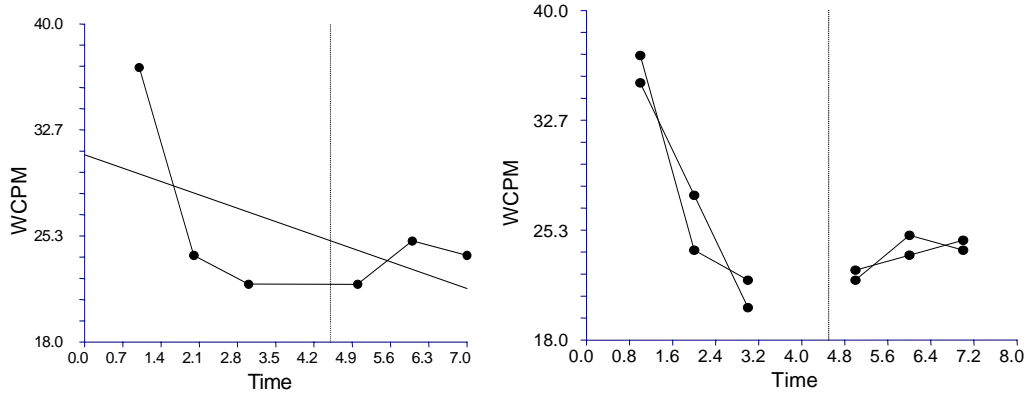
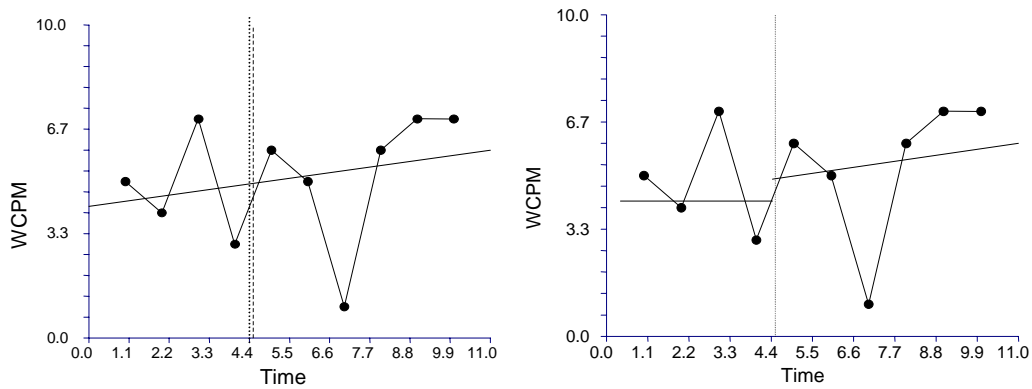
Examiner says:

“Some students your age can read this passage very well, while other students your age can’t read this passage at all. In this row of numbers, 1 means students your age at your home school who cannot read the passage at all, 3 means most students your age at your home school, and 5 means the very best reader your age at your home school (point to each number as you describe its meaning; If the student does not appear to understand the scale, you may explain further). **Circle** the number that you think shows how well you can read this passage compared to other students your age at your home school.”

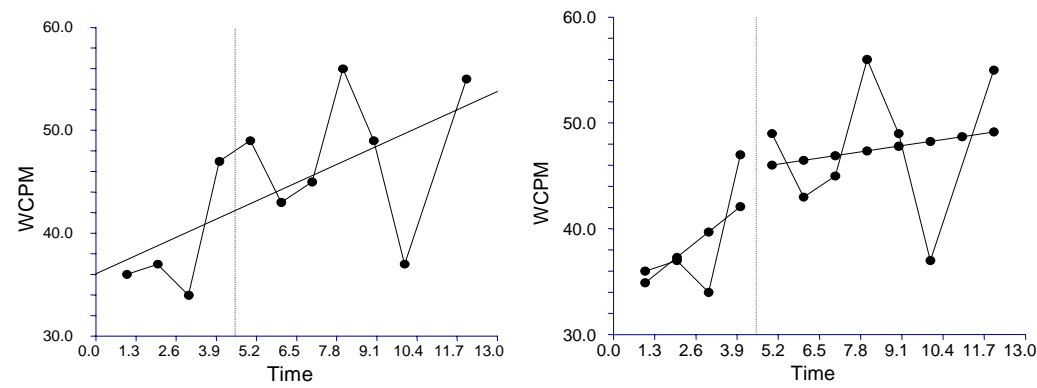
1 2 3 4 5

APPENDIX B

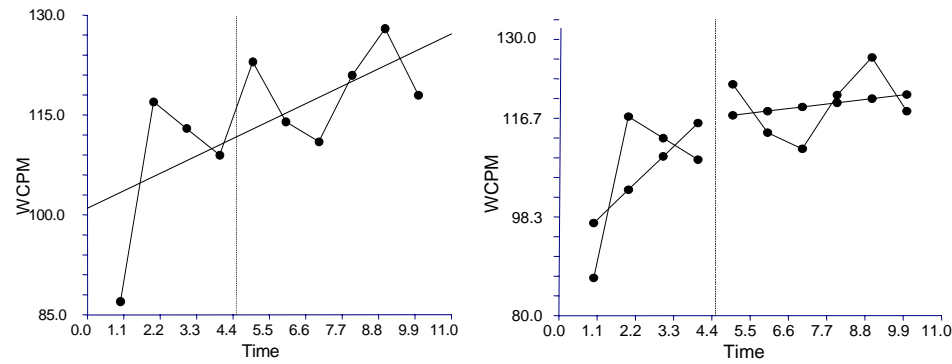
GRAPHS OF RAW DATA (LEFT) AND MEAN PLUS TREND DATA (RIGHT) FOR
FLUENCY SCORES OF EACH STUDENT

A**B**

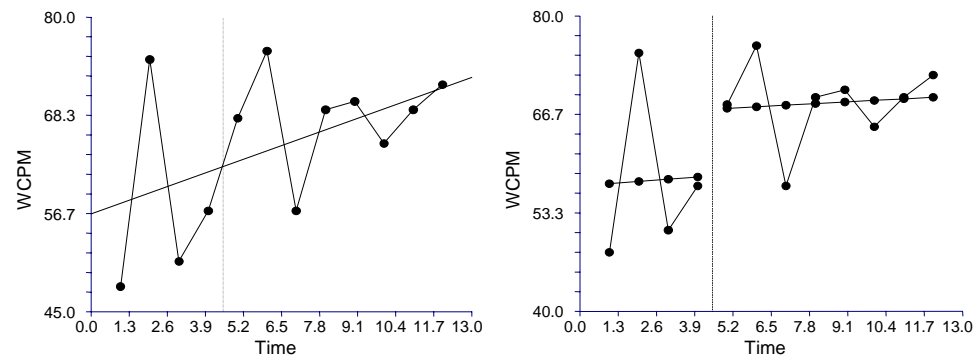
C

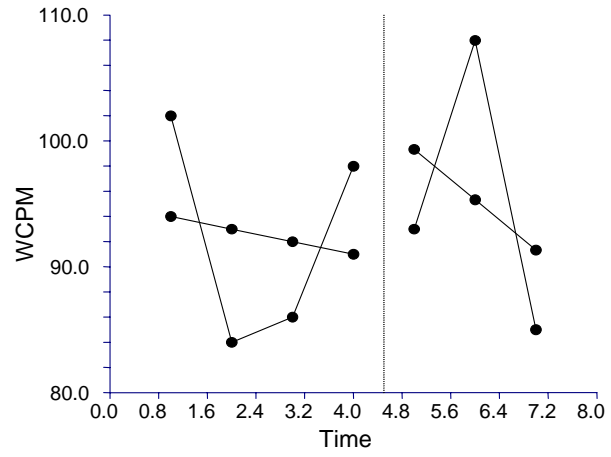
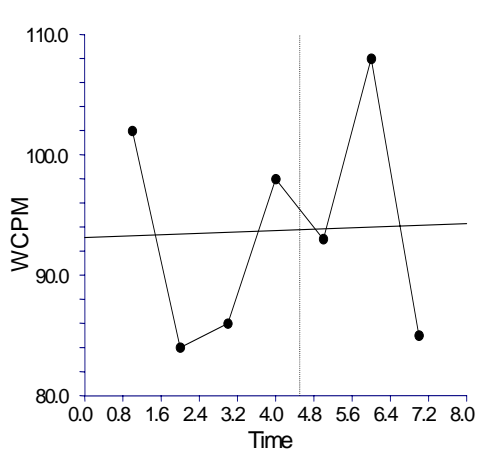
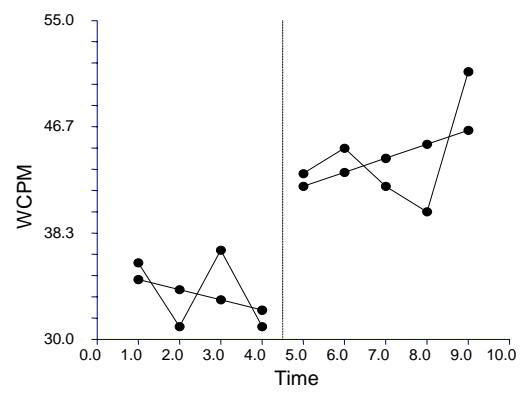
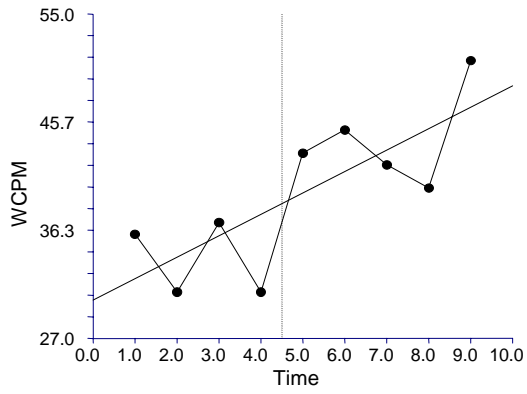
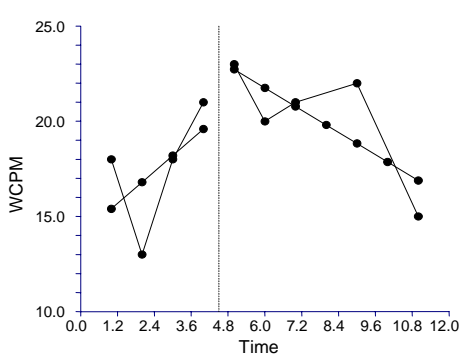
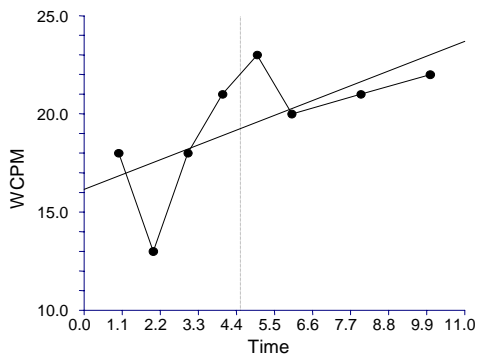


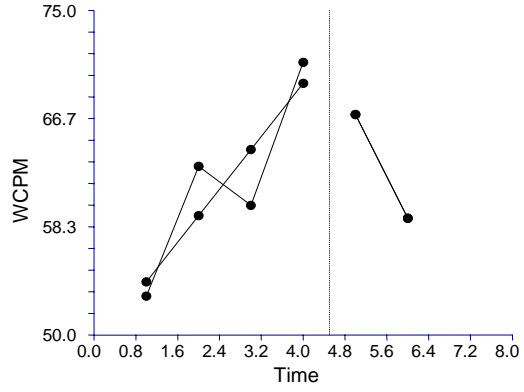
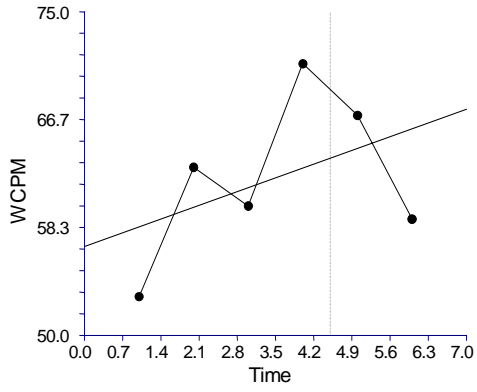
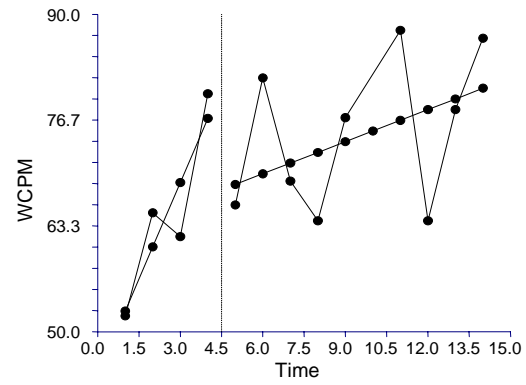
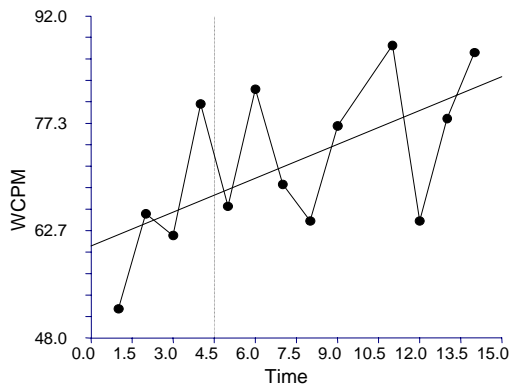
D

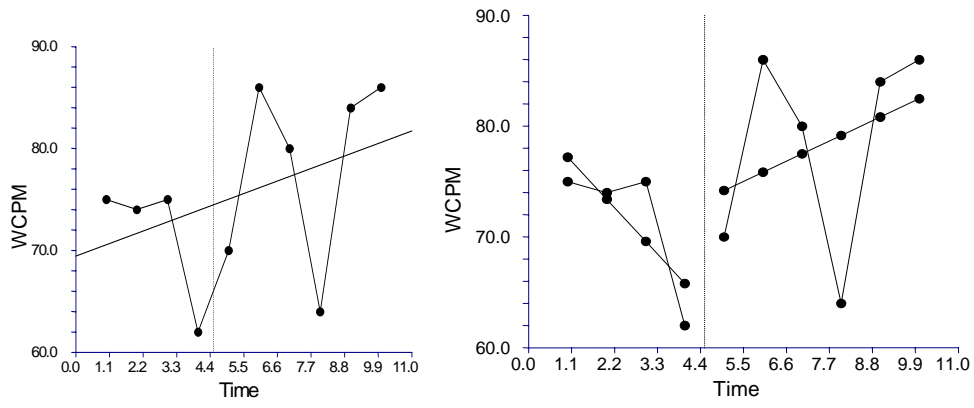
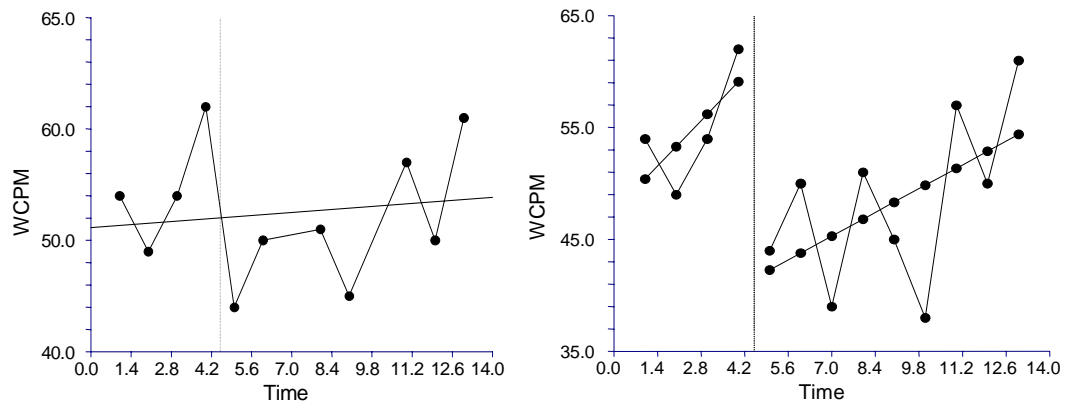
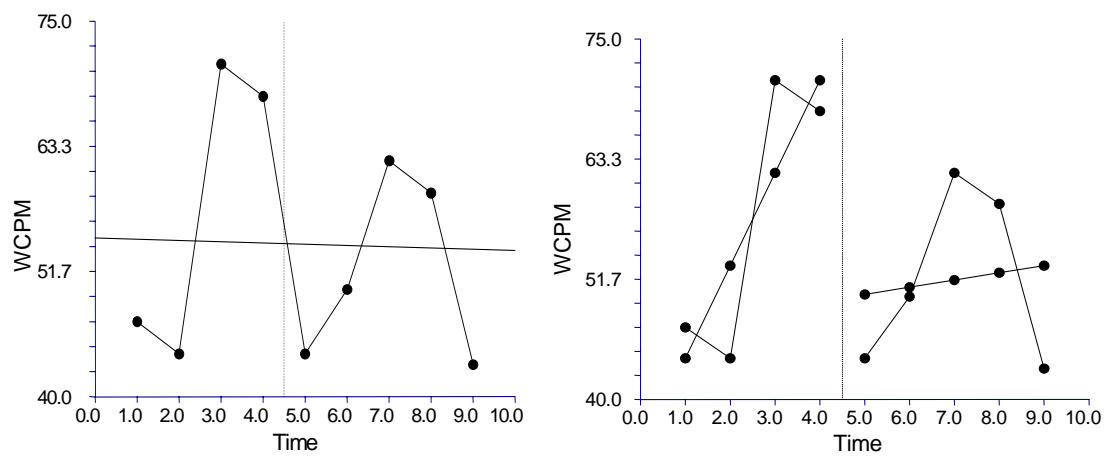


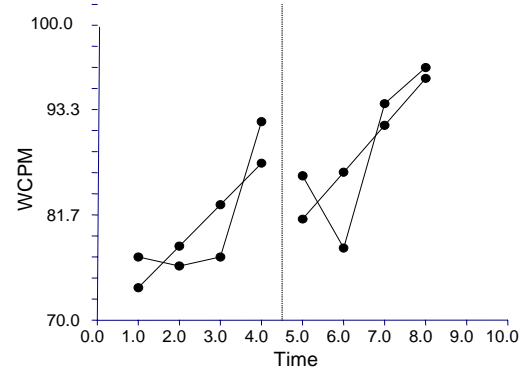
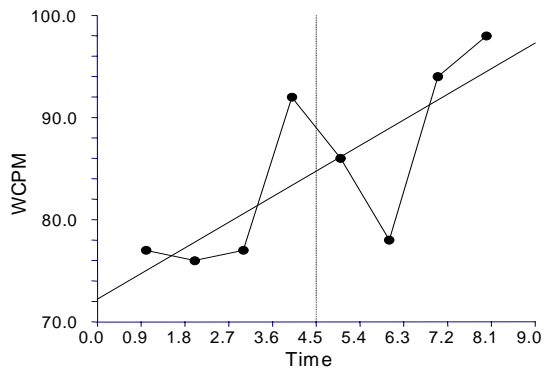
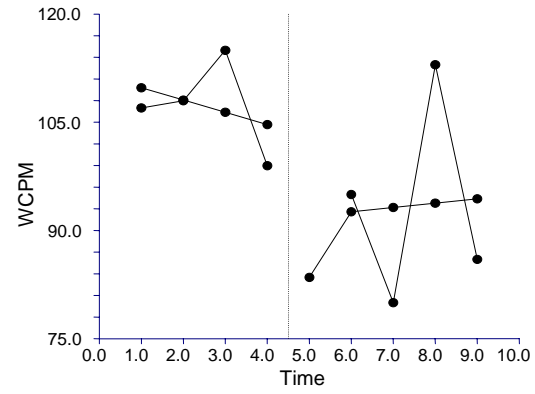
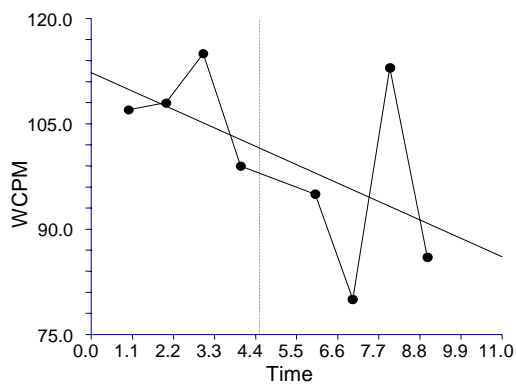
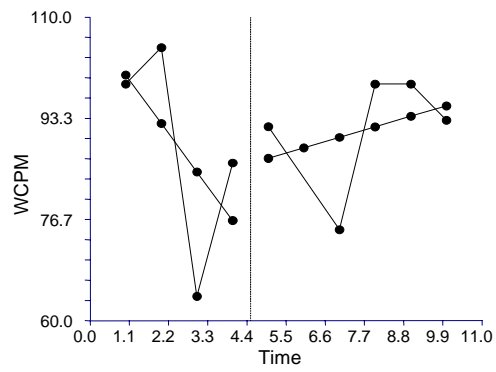
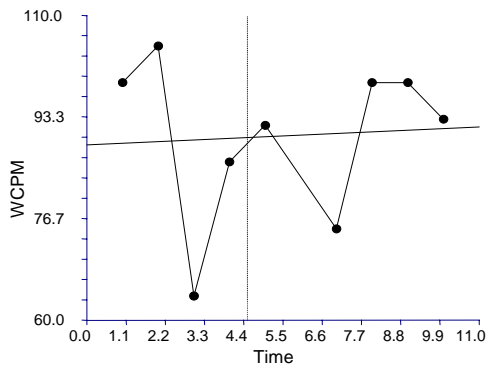
E

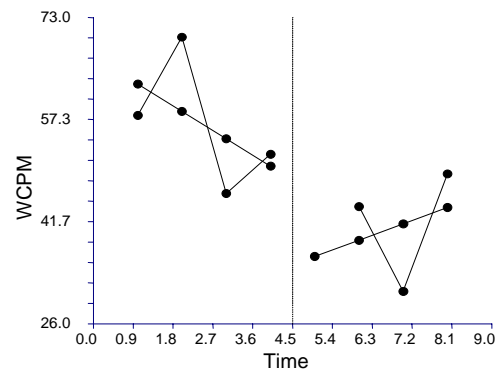
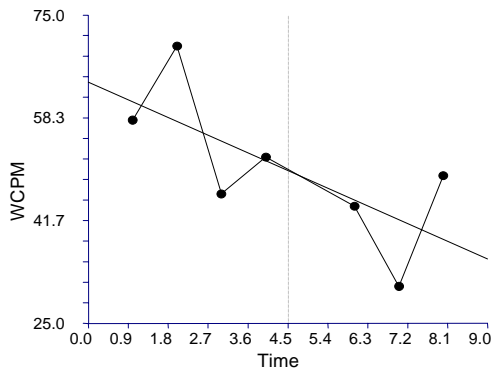
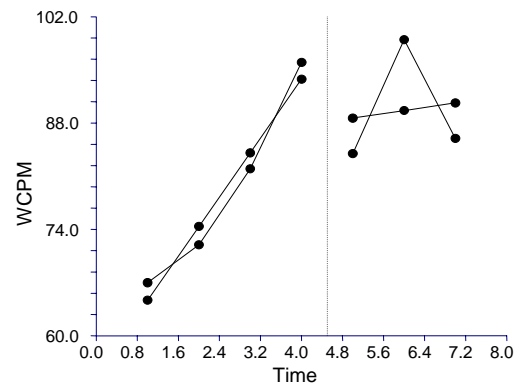
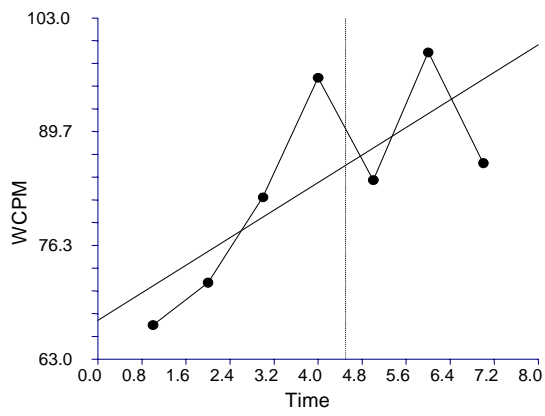
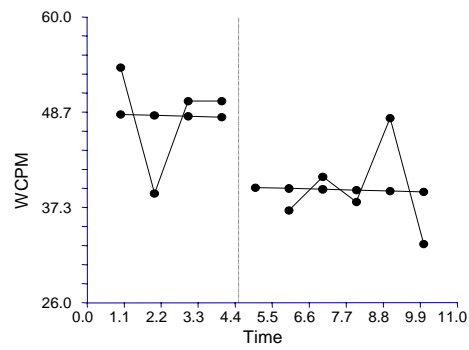
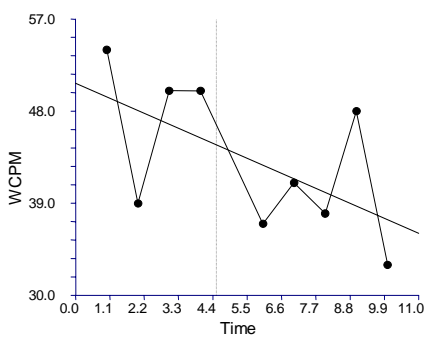


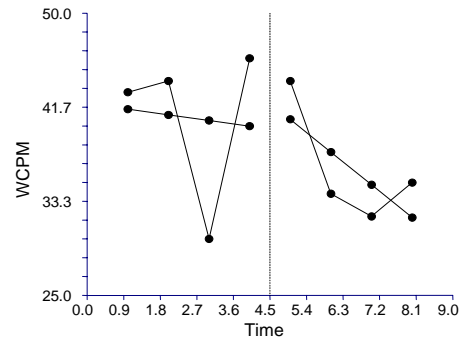
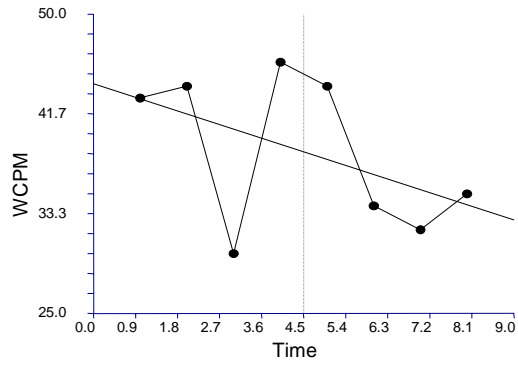
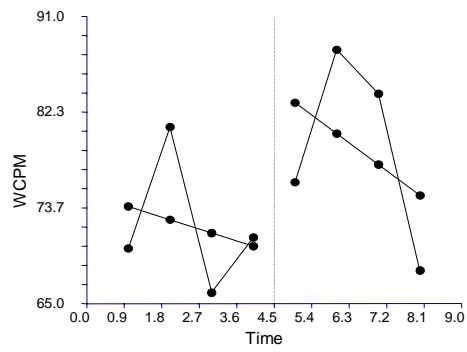
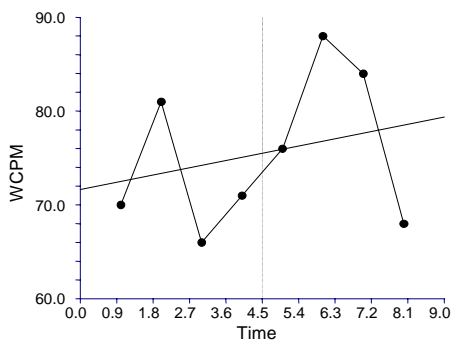
F**G****H**

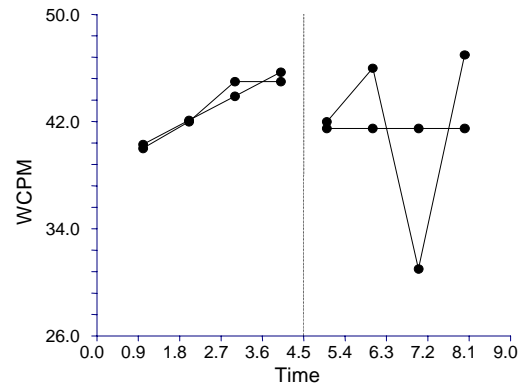
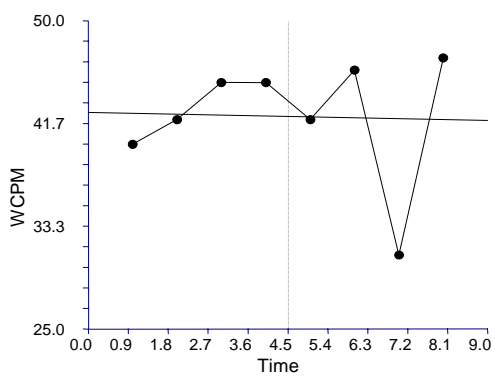
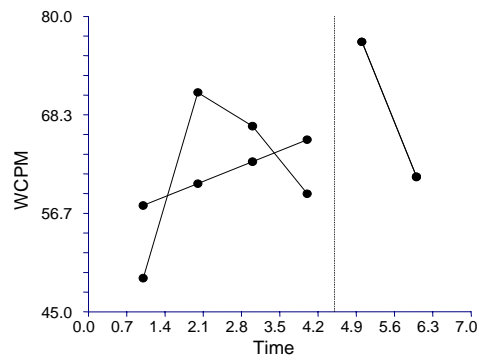
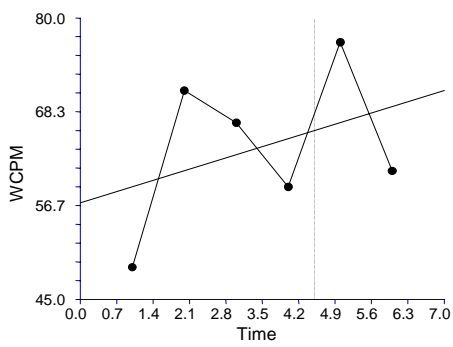
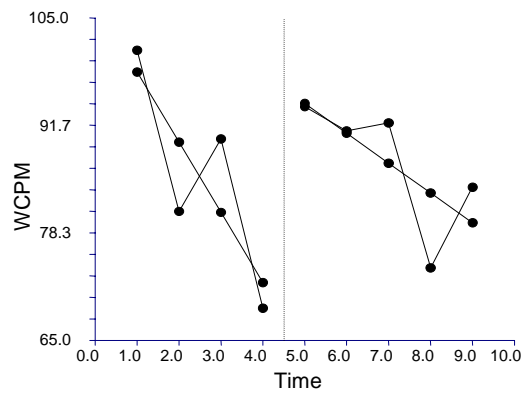
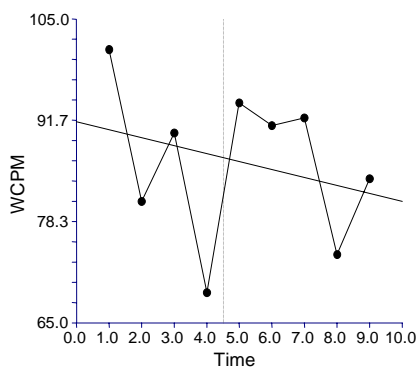
I**J**

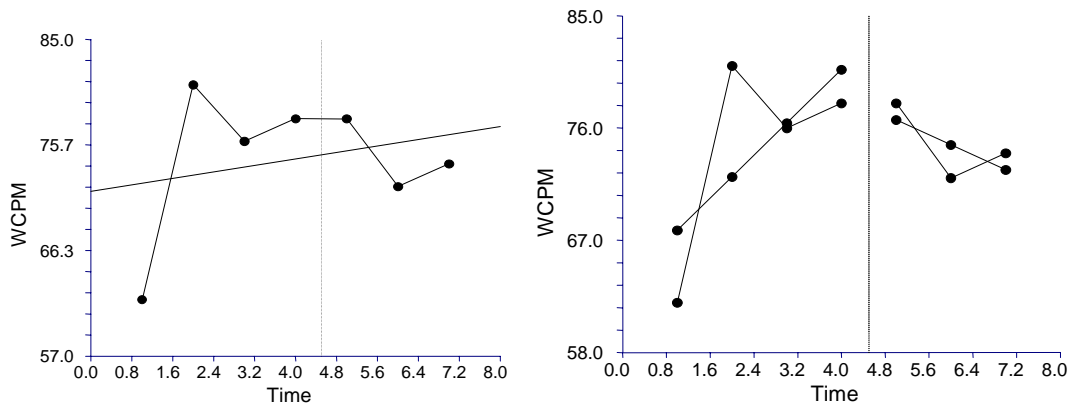
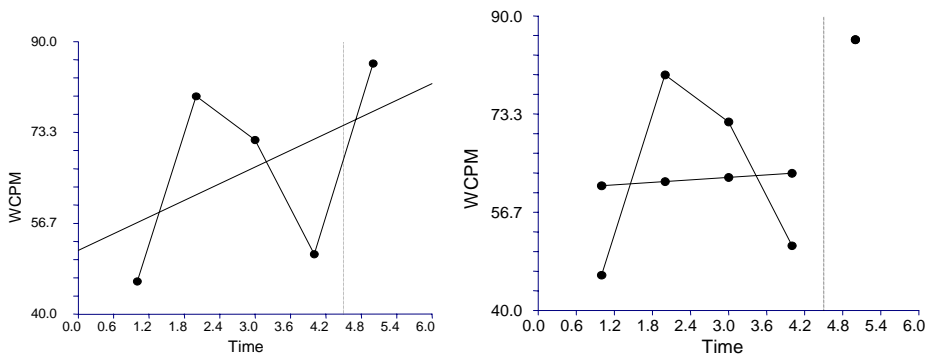
K**L****M**

N**O****P**

Q**R****S**

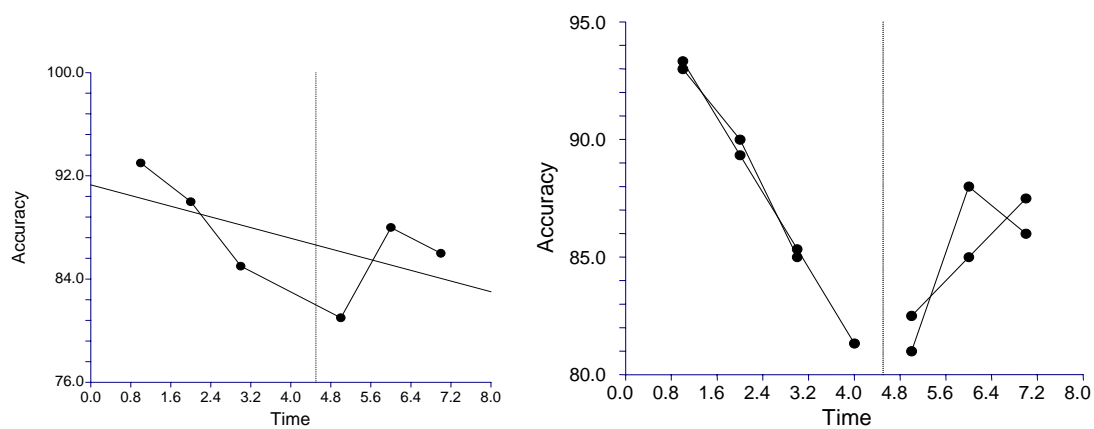
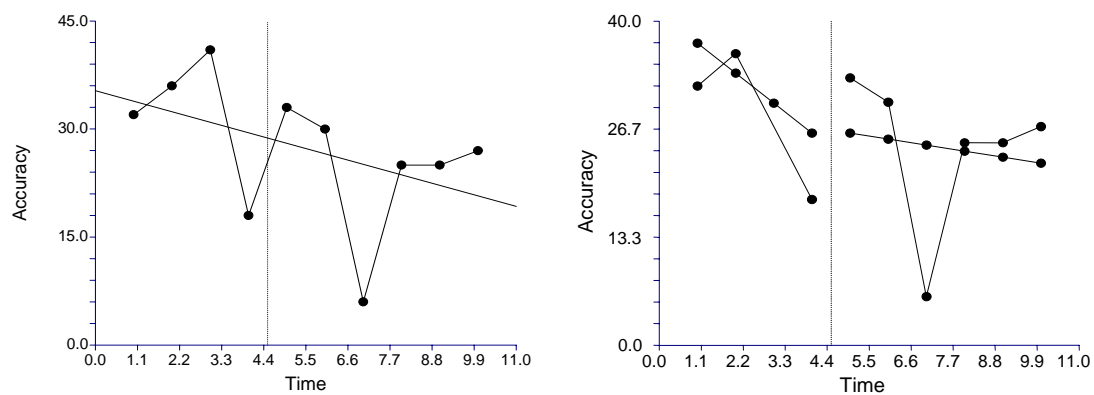
T**U**

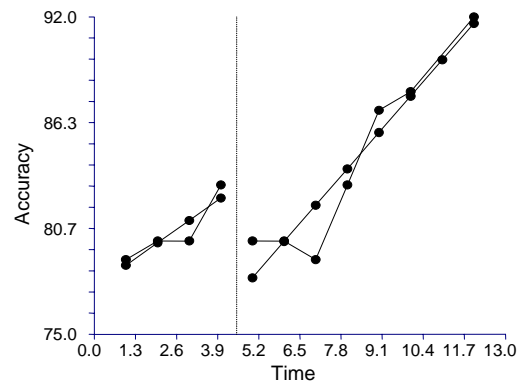
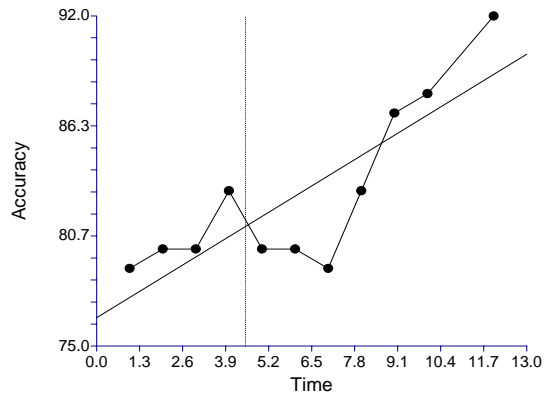
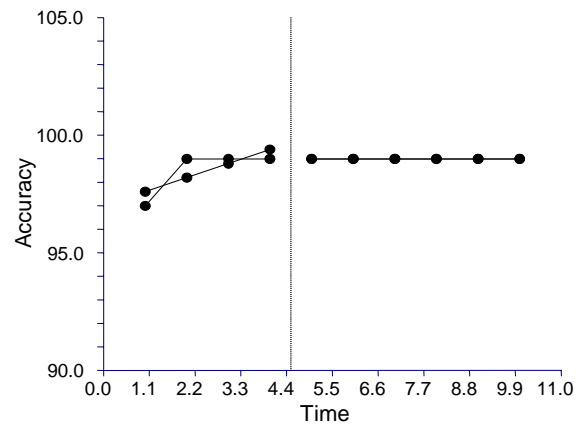
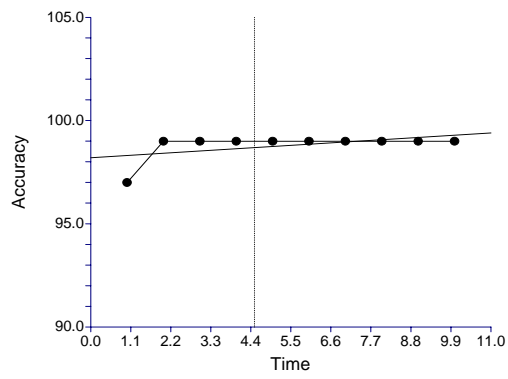
V**W****X**

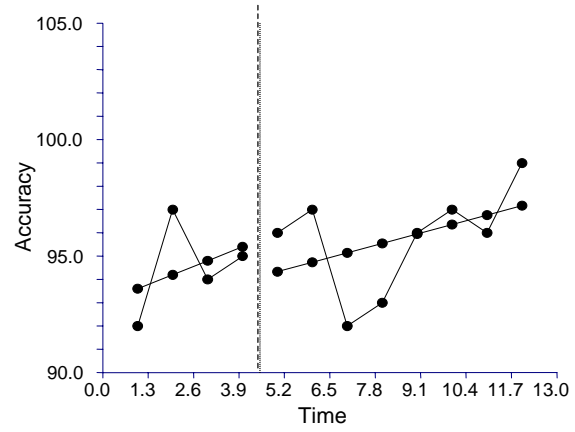
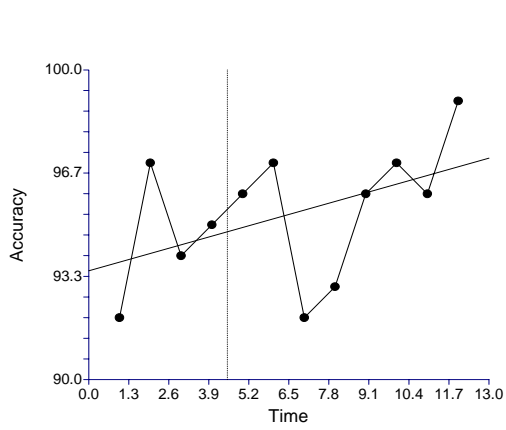
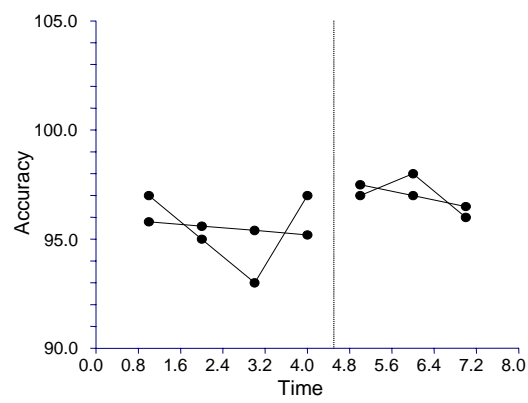
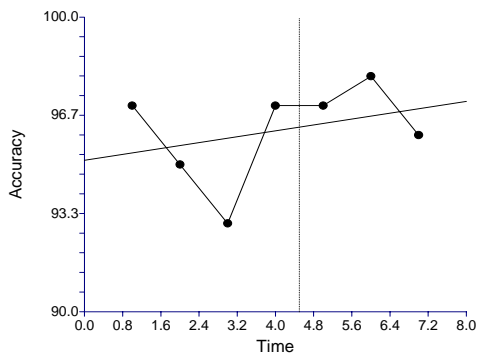
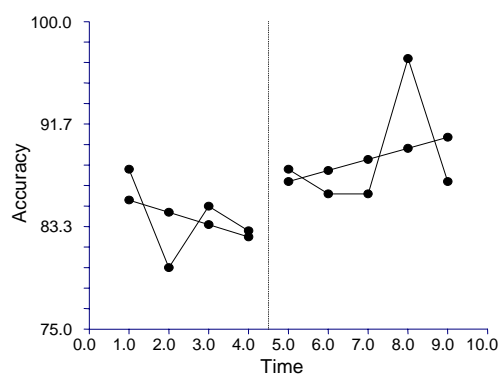
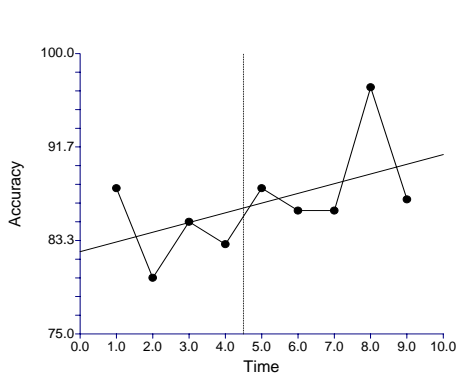
Y**Z**

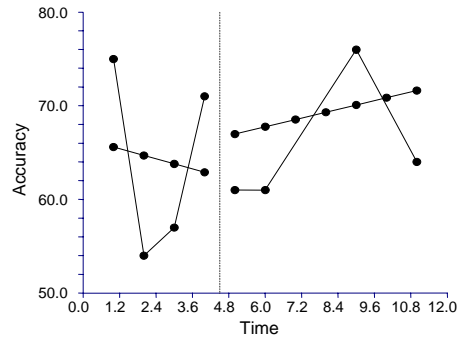
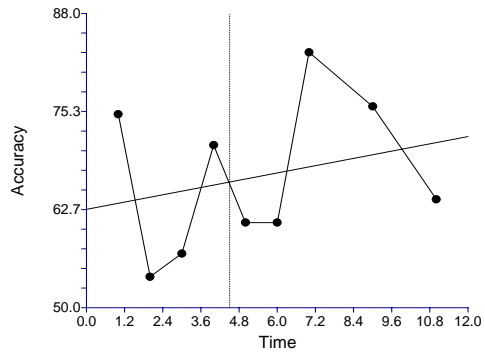
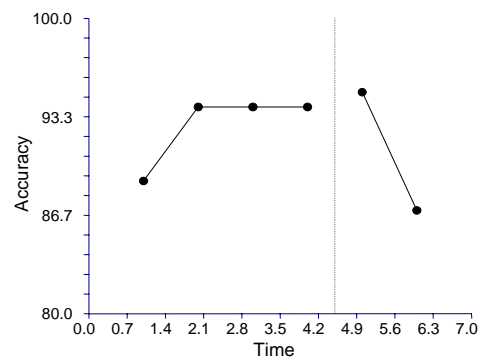
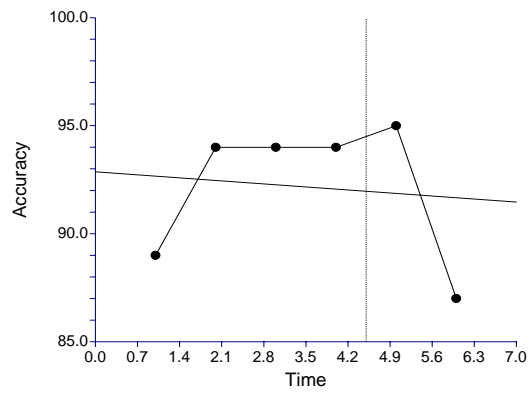
APPENDIX C

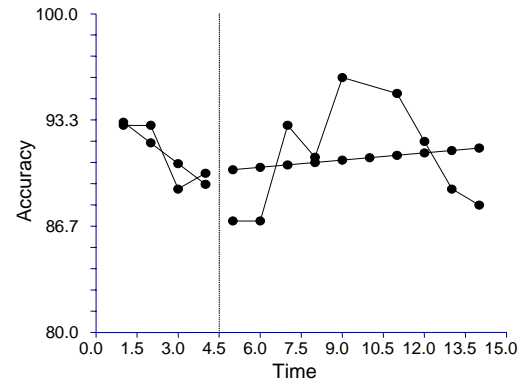
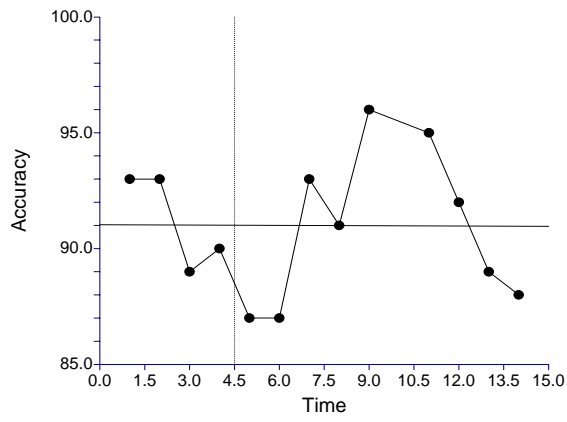
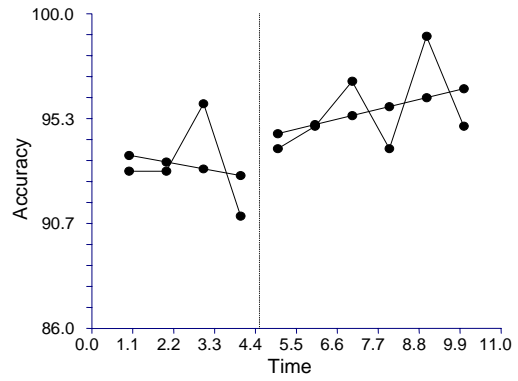
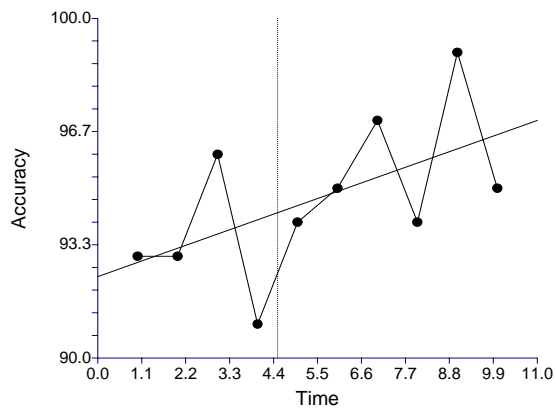
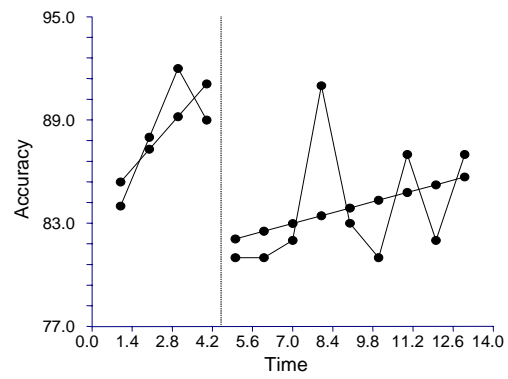
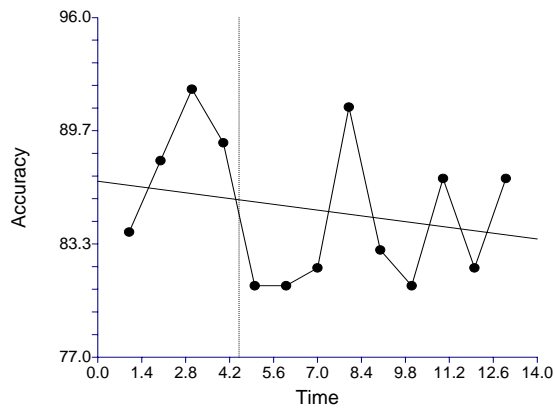
GRAPHS OF RAW DATA (LEFT) AND MEAN PLUS TREND DATA (RIGHT) FOR
ACCURACY SCORES OF EACH STUDENT

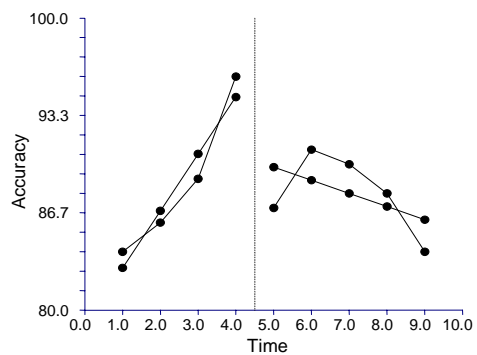
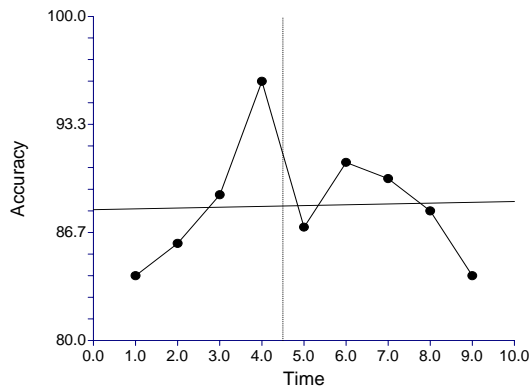
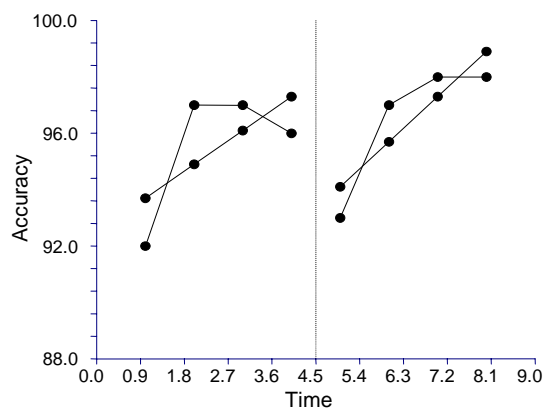
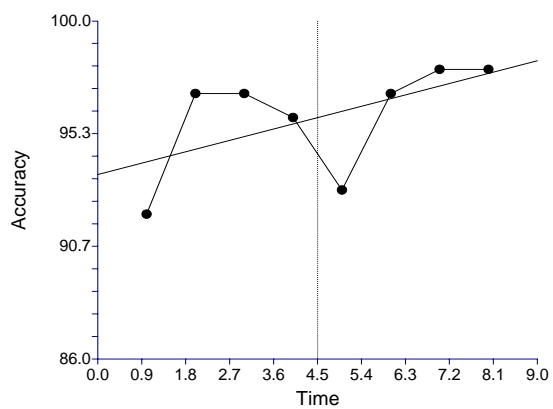
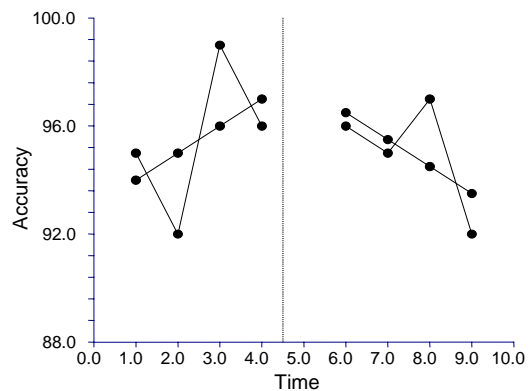
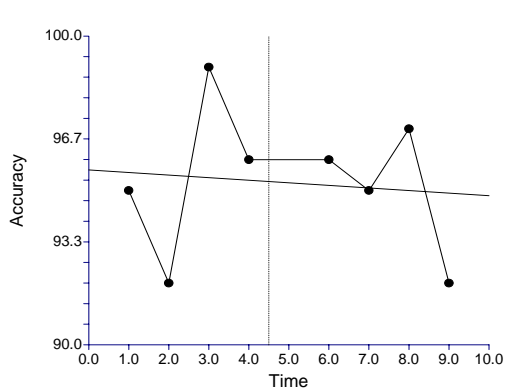
A**B**

C**D**

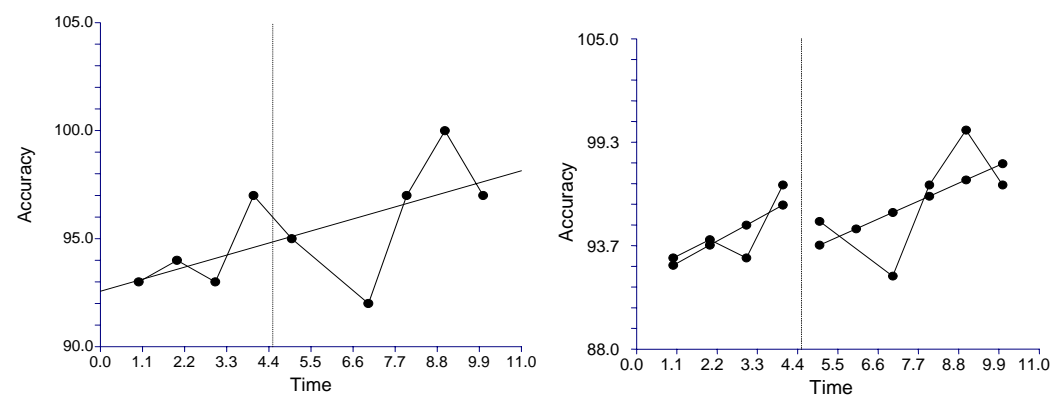
E**F****G**

H**I**

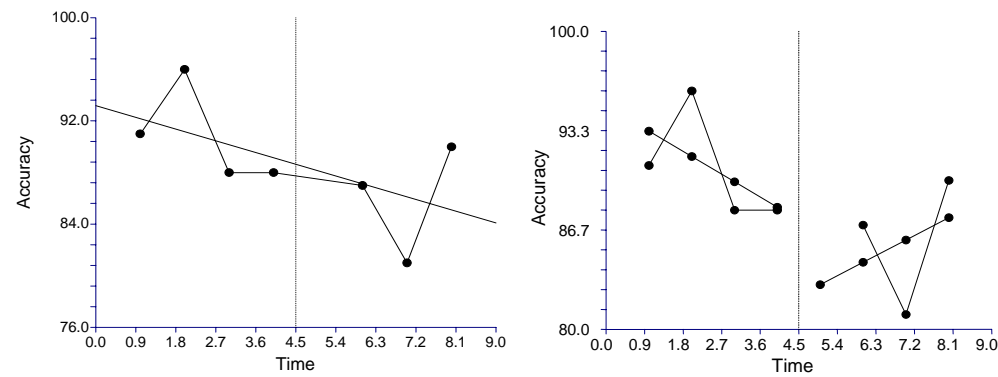
J**K****L**

M**N****O**

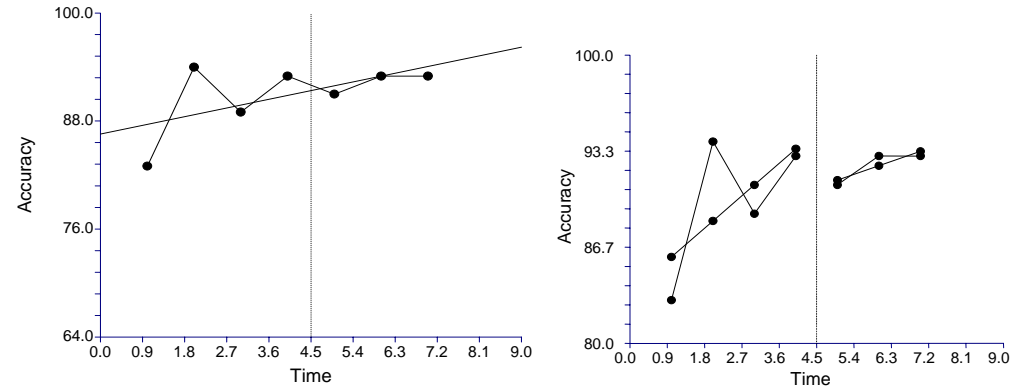
P

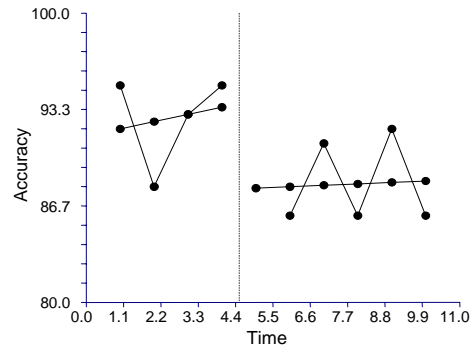
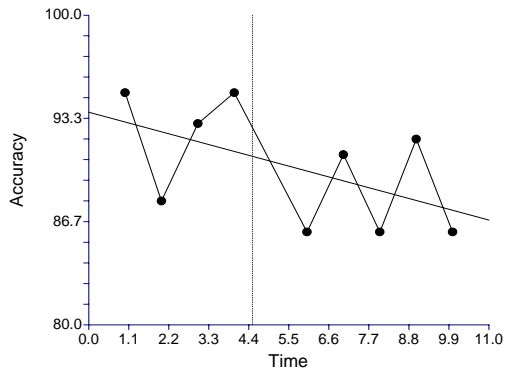
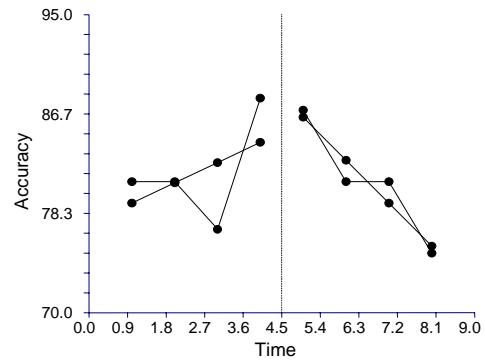
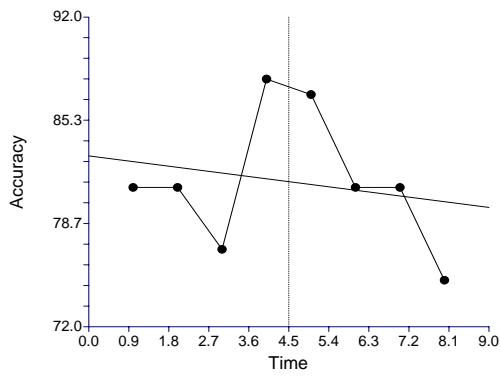


Q

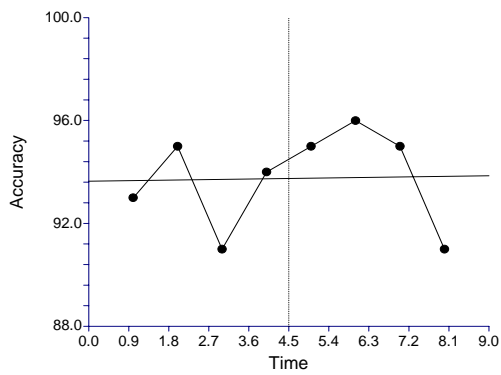


R

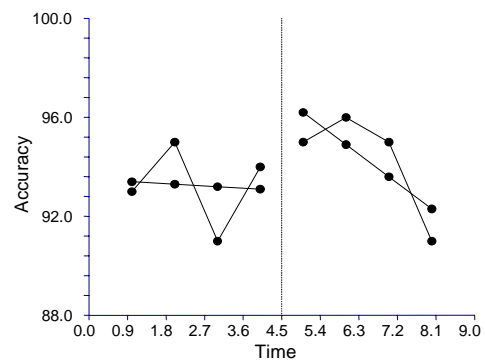


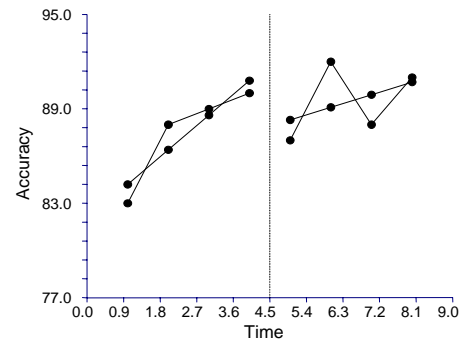
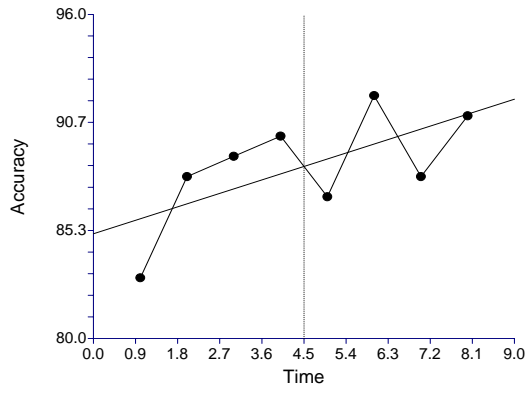
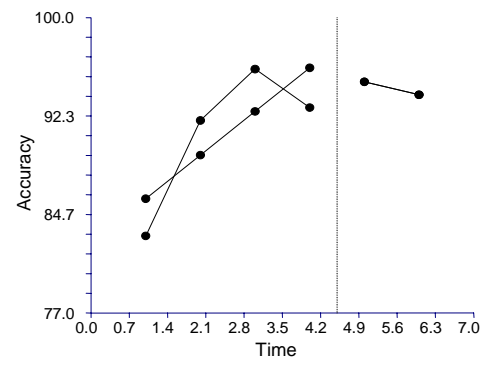
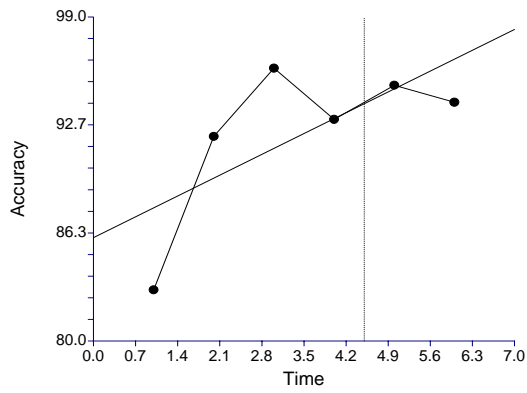
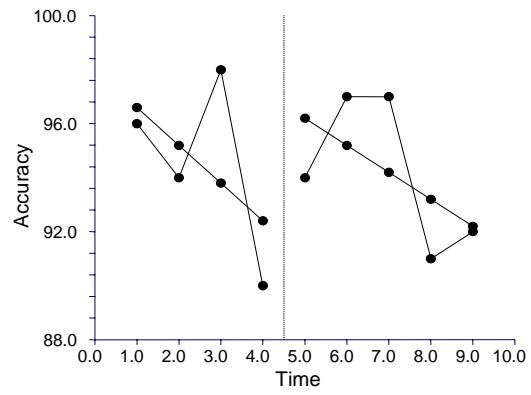
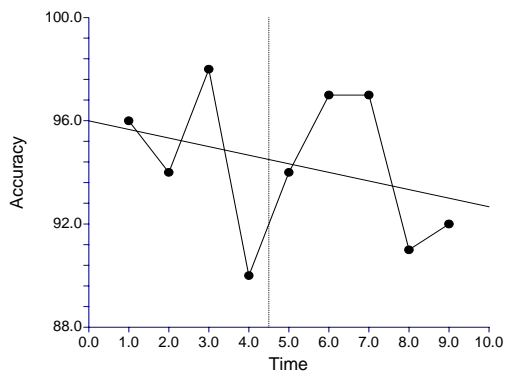
S**T****U**

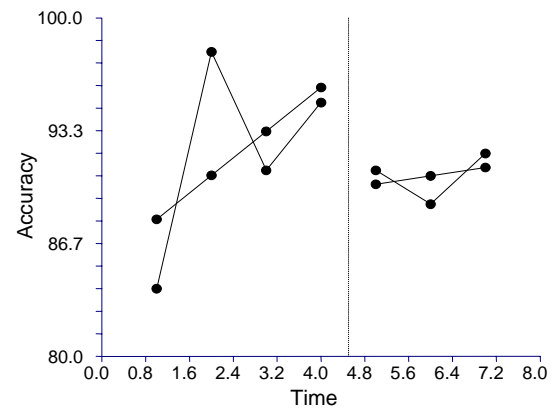
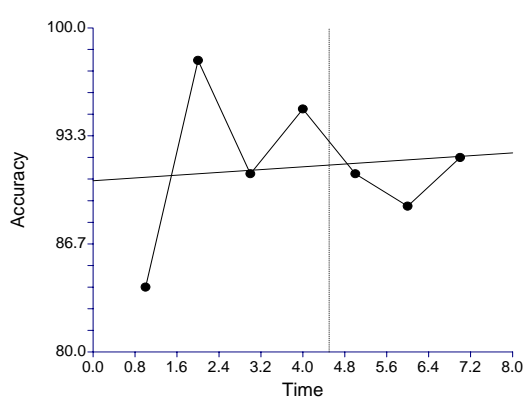
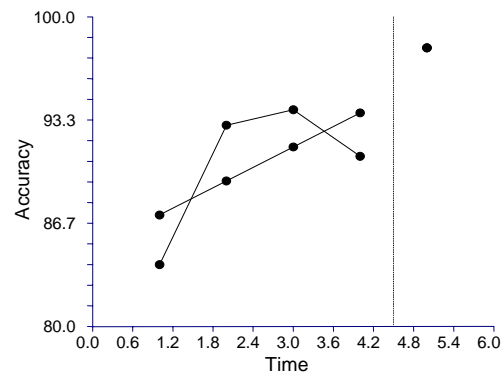
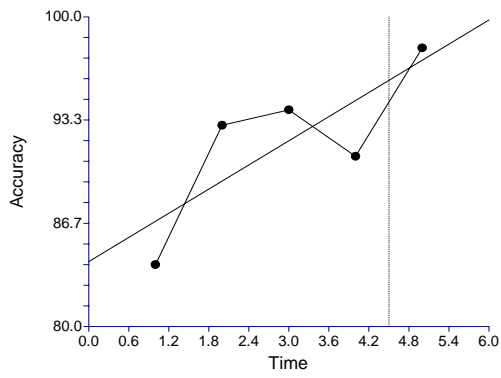
Student 21



Student 22



VWX

Y**Z**

VITA

Elizabeth Bubonic Olson

1528 Wynfield Drive
Little Elm, TX 75068
(972) 294-1672
txhawk29@hotmail.com

Educational Background

Ph.D., School Psychology, Texas A&M University.

B.S., with Honors in Psychology, University of Iowa. December, 1998.

Professional Experience

2003-2004	Internship in School Psychology, Lewisville ISD, Lewisville, Texas
2001-2003	Graduate Assistant, College Station, Texas, Project Coordinator Internet-based training for Master Reading Teachers
2000-2003	Graduate Assistant, College Station, Texas, developing Internet-based training for Educational Diagnosticians and Master Reading Teachers
1999-2000	Graduate Assistant, College Station, Texas—Dr. Richard Parker
1999-2001	Examiner for Texas Instruments research grant at Head Start school in Dallas, Texas
1999-2000	Research assistant for study on Curriculum-based Measurement with Bilingual elementary school students—Carolyn Denton
1999-2000	Data collector for Texas Primary Reading Inventory pilot study